CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Bradford Scholars

UNIVERSITY of BRADFORD

Library

University of Bradford eThesis

This thesis is hosted in Bradford Scholars – The University of Bradford Open Access repository. Visit the repository for full metadata or to contact the repository team

© University of Bradford. This work is licenced for reuse under a Creative Commons Licence.

Performance Modelling and Evaluation of Heterogeneous Wired/Wireless Networks under Bursty Traffic

Yulei WU

Ph.D

2010

Performance Modelling and Evaluation of Heterogeneous Wired/Wireless Networks under Bursty Traffic

Analytical models for performance analysis of communication networks in multi-computer systems, multi-cluster systems, and integrated wireless systems

Yulei WU

Submitted for the degree of Doctor of Philosophy

School of Computing, Informatics and Media

University of Bradford

2010

Abstract

Computer networks can be classified into two broad categories: wired networks and wireless networks, according to the hardware and software technologies used to interconnect the individual devices. Wired interconnection networks are hardware fabrics supporting communications between individual processors in high-performance computing systems (e.g., multi-computer systems and cluster systems). On the other hand, due to the rapid development of wireless technologies, wireless networks have emerged and become an indispensable part for people's lives. The integration of different wireless technologies is an effective approach to accommodate the increasing demand of the users to communicate with each other and access the Internet.

This thesis aims to investigate the performance of wired interconnection networks and integrated wireless networks under the realistic working conditions. Traffic patterns have a significant impact on network performance. A number of recent measurement studies have convincingly demonstrated that the traffic generated by many real-world applications in communication networks exhibits bursty arrival nature and the message destinations are non-uniformly distributed. Analytical models for the performance evaluation of wired interconnection networks and integrated wireless networks have been widely reported. However, most of these models are developed under the simplified assumption of non-bursty Poisson process

with uniformly distributed message destinations.

To fill this gap, this thesis first presents an analytical model to investigate the performance of wired interconnection networks in multi-computer systems. Secondly, the analytical models for wired interconnection networks in multi-cluster systems are developed. Finally, this thesis proposes analytical models to evaluate the end-to-end delay and throughput of integrated wireless local area networks and wireless mesh networks. These models are derived when the networks are subject to bursty traffic with non-uniformly distributed message destinations which can capture the burstiness of real-world network traffic in the both temporal domain and spatial domain. Extensive simulation experiments are conducted to validate the accuracy of the analytical models. The models are then used as practical and cost-effective tools to investigate the performance of heterogeneous wired or wireless networks under the traffic patterns exhibited by real-world applications.

Keywords: interconnection networks, multi-computer systems, multi-cluster systems, integrated wireless networks, bursty traffic, non-uniform traffic, analytical modelling

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACN	IntrA-Communication Network
C/D	Concentrator/Dispatcher
CTS	Clear to send
COTS	Commodity-off-the-shelf
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
DCF	Distributed coordination function
ECN	IntEr-Communication Network
EDCA	Enhanced distributed channel access
HPC	High-performance computing
IBA	InfiniBand architecture
ICN	Integrated Communication Network
IN	Interconnection network
LM	Local memory
MAC	Medium access control
MC	Mesh client
MIN	Multistage interconnection network
MMPP	Markov-modulated Poisson process
MR	Mesh router
NoC	Network-on-chip
NCA	Nearest common ancestor
OMNeT++	Objective Modular Network Testbed in C++

PE	Processing element
PHY	Physical
QoS	Quality-of-service
RTS	Request to send
SCV	Squared coefficient of variation
SoC	System-on-chip
WLAN	Wireless local area network
WMAN	Wireless metropolitan area network
WMN	Wireless mesh network
WPAN	Wireless personal area network
WSN	Wireless sensor network

Acknowledgements

First and foremost I would like to express my sincere gratitude to my research supervisor Dr. Geyong Min, not only for his continuous guidance and close collaboration in his role as my advisor, but also for the supportive way he has helped me through the difficult moments of my research work. His suggestions and constructive feedback guided me through the work on this thesis.

Special thanks to my parents for their endless source of support, love, and encouragement during my research. I am indebted to my wife for putting up with many husbandless evenings and weekends when I was doing my research. I would also like to thank my friends and colleagues for their emotional support when it was most needed.

Lastly, but definitely not least, I would like to take this opportunity to thank the Department of Computing, School of Computing, Informatics and Media, University of Bradford for providing me the facilities to carry out my research work.

List of Publications

Y. Wu, G. Min, M. Ould-Khaoua, and D.D. Kouvatsos, "An Analytical Model for Hypercubic Networks under Correlated Traffic with Non-Uniform Destination Distributions," *Traffic and Performance Engineering for Heterogeneous Networks*, D.D. Kouvatsos (ed.), River Publishers, Gistrup, Denmark, ISBN: 978-87-92329-16-5, 2009.

Y. Wu, G. Min, M. Ould-Khaoua, H. Yin, and L. Wang, "Analytical Modelling of Networks in Multicomputer Systems under Bursty and Batch Arrival Traffic," *The Journal of Supercomputing*, vol. 51, no. 2, pp. 115-130, 2010.

Y. Wu, G. Min, M. Ould-Khaoua, and H. Yin, "An Analytical Model for Torus Networks in the Presence of Batch Message Arrivals with Hot-spot Destinations," *International Journal of Automation and Computing*, vol. 6, no. 1, pp. 38-47, 2009.

G. Min, Y. Wu, K. Li, and A.Y. Al-Dubai, "Performance Modelling and Optimisation of Integrated Wireless LANs and Mesh Networks," *International Journal of Communication Systems*, in press, online version available at doi: 10.1002/dac.1074, 2009.

Y. Wu and G. Min, "Delay and Throughput Analysis in Multi-Hop Cognitive Radio Networks," *IEEE ICC'10*, 2010.

Y. Wu and G. Min, "An Analytical Model for Hybrid Wireless Networks under Bursty and Correlated Traffic," *IEEE ICC'10*, 2010.

G. Min and Y. Wu, "A Performance Model for Multi-Hop Cognitive Mesh Networks," *IEEE ICC'10*, 2010.

G. Min, Y. Wu, K. Li, and B. Javadi, " A Performance Model for Multi-Cluster Computing Systems in the Presence of Hot-Spot Traffic," *IEEE ICC'10*, 2010.

Y. Wu, G. Min, K. Li, and A. Y. Al-Dubai, "A Performance Model for Integrated

Wireless Mesh Networks and WLANs with Heterogeneous Stations," *IEEE GLOBECOM'09*, pp. 1-6, 2009.

Y. Wu, G. Min, K. Li, and B. Javadi, "Performance Analysis of Communication Networks in Multi-Cluster Systems under Bursty Traffic with Communication Locality," *IEEE GLOBECOM'09*, pp. 1-6, 2009.

G. Min, Y. Wu, M. Ould-Khaoua, H. Yin, and K. Li, "Performance Modelling and Analysis of Interconnection Networks with Spatio-Temporal Bursty Traffic," *IEEE GLOBECOM'09*, pp. 1-6, 2009.

G. Min, Y. Wu, and L.T. Yang, "Performance Modelling and Analysis of Integrated WLANs and Internet-Access Mesh Networks," *IEEE CSE'09*, pp. 1-8, 2009.

Y. Wu, G. Min, K. Li, and B. Javadi, "Performance Modelling of Large-Scale Multi-Cluster Computing Systems under Batch Arrival Traffic," *ICAC'09*, pp. 23-29, 2009.

Y. Wu, G. Min, G. Wang, and J. Jiang, "Modelling of Heterogeneous Wireless Networks under Batch Arrival Traffic with Communication Locality," *IEEE WCNC'09*, pp. 1-6, 2009.

G. Min, Y. Wu, K. Li, and A. Y. Al-Dubai, "Performance Analysis of Two-Tier Wireless Mesh Networks for Achieving Delay Minimisation," *IEEE WCNC'09*, pp. 1-6, 2009.

Y. Wu, G. Min, M. Ould-Khaoua, and H. Yin, "Analytical Modelling of Pipelined Circuit Switching with Bursty and Hot-Spot Traffic," *IEEE HPCC'08*, pp. 470-477, 2008.

Y. Wu, G. Min, M. Ould-Khaoua, and H. Yin, "Modelling and Performance Analysis of Torus Networks in the Presence of Bursty Traffic and Hot-Spot Destinations," *ICAC'08*, pp. 180-185, 2008.

Y. Wu, G. Min, L. Wang, "Performance Analysis of Interconnection Networks Under Bursty and Batch Arrival Traffic," *ICA3PP'07*, pp. 25-36, 2007.

Y. Wu, G. Min, M. Ould-Khaoua, and H. Yin, "An Analytical Model for Torus Networks in the Presence of Hot-Spot Traffic with Batch Arrivals," *IEEE MASCOTS'07*, pp. 395-401, 2007.

G. Min, Y. Wu, L. Wang, and M. Ould-Khaoua, "Performance Modelling of Adaptive Routing in Hypercubic Networks under Non-Uniform and Batch Arrival Traffic," *IEEE LCN'07*, pp. 583-590, 2007.

Chapter 1

Introduction

Computer networks which have existed for more than thirty years facilitate communications among devices connected to the network. Computer networks can be classified into two broad categories: wired networks and wireless networks, according to the hardware and software technologies that are used to interconnect the individual devices in the network. The devices in the wired networks are connected by physical wires, including optical fiber, Ethernet, etc., and communicate with each other through the wired lines. On the other hand, the devices use radio waves or infrared signals as a transmission medium in wireless networks.

1.1 Motivation

Performance modelling and evaluation of heterogeneous wired or wireless networks have received considerable research attentions. Although the wired interconnection networks (INs) have attracted numerous research efforts, it is still required to develop innovative performance tools and methods to keep up with the rapid evolution and ever increasing complexity of high-performance computing (HPC) systems, e.g., multi-computer systems and multi-cluster systems.

A *multi-computer system* is composed of hundreds or thousands of processors, which exchange data and synchronise with each other by passing messages through a multi-hop IN [17]. Improvements in computational and communication technologies have made it economically feasible to conglomerate multiple clusters of commodity-off-the-shelf (COTS) networked components, known as *multi-cluster systems* [58, 95]. Due to the undeniable cost-effectiveness, multi-cluster systems have been gaining wide acceptance for solving many complex scientific, engineering, and commercial computation applications [1, 32]. Such systems are embracing the environment with the even greater heterogeneity of computation and communication capacities in individual clusters [115]. The underlying IN is one of the key components that dominate the performance of these systems [23].

On the other hand, due to the rapid development of wireless technologies, wireless networks have become an indispensable part for people's lives. The technologies for *wireless networks*, such as cellular networks, Wi-Fi, WiMAX, wireless sensor networks (WSNs), wireless personal area networks (WPANs), wireless local area networks (WLANs), and wireless metropolitan area networks (WMANs) have evolved quickly and emerged with the success of wireless communications [112]. The integration of different wireless technologies becomes an effective approach to accommodate the increasing demand of the users to communicate with each other and access the Internet [4, 87].

A number of measurement studies have convincingly demonstrated that the traffic generated by many real-world applications in wired INs and wireless networks exhibits bursty nature [19, 100, 110, 111, 125] and the message destinations are non-uniformly distributed over the network nodes [2, 15, 68, 93]. However, most existing studies of wired INs [5, 21, 56, 73, 91] and wireless networks [7, 50, 51, 76] are

based on the unrealistic assumption of non-bursty Poisson arrivals with uniformly distributed message destinations which are not suitable for evaluation and optimisation of network performance. Moreover, the existing performance studies of wired INs in multi-cluster systems have mainly focused on the homogeneous clusters and the evaluations are confined to a single cluster only [22, 59], which cannot capture the heterogeneity of computation and communication capacities in individual clusters.

In addition, most of studies reported in the current literature [7, 8, 50, 51, 66, 71, 76] have focused on the performance evaluation of a single wireless network. Although several models for integrated wireless networks have been presented [87], the quality-of-service (QoS) performance measures including delay and throughput are not addressed. Furthermore, there is hardly any model reported in the literature to handle the bursty traffic with non-uniformly distributed message destinations in integrated wireless networks.

1.2 Research aims

The central theme of this thesis is to develop innovative and cost-effective tools to evaluate the performance of wired INs and integrated wireless networks.

The main aims of the research are outlined as follows:

- to develop performance models for wired INs in multi-computer systems under the traffic patterns exhibited by real-world applications;
- to propose analytical models for the performance evaluation of wired INs in heterogeneous multi-cluster systems under realistic traffic patterns;
- to develop performance models to investigate the QoS performance measures including end-to-end delay and throughput in integrated wireless networks (e.g.,

WLANs, WSNs, etc.) and wireless mesh networks (WMNs) under the traffic patterns exhibited by real-world applications;

- to use these models as practical and cost-effective tools to evaluate and optimise the performance of wired INs and integrated wireless networks.

1.3 Contributions

The principal contributions of this thesis are:

- New analytical models are developed to investigate the performance of wired INs in multi-computer systems in the presence of bursty traffic with non-uniformly distributed message destinations which can capture the burstiness of real-world network traffic in the both temporal domain and spatial domain. These models and tools are validated through extensive simulation experiments of an actual system;
- Original analytical models are proposed to evaluate the performance of wired INs in heterogeneous multi-cluster systems under bursty traffic with non-uniformly distributed message destinations. The validity and accuracy of the model are demonstrated by comparing analytical results to those obtained through extensive simulation experiments;
- New analytical models are developed to investigate the QoS performance measures, e.g., end-to-end delay and throughput, for integrated WLANs and WMNs in the presence of bursty traffic with non-uniformly distributed message destinations. Extensive simulation experiments are conducted to validate the accuracy of the analytical model;
- The developed analytical models are used as practical and cost-effective tools to investigate the performance of wired INs and integrated wireless networks.

1.4 Thesis organisation

The remainder of this thesis is organised as follows.

Chapter 2 presents the background research and related work of wired INs and integrated wireless networks. The models which can characterise the nature of realistic network traffic exhibited by real-world applications are illustrated. The methodologies for the performance study of wired INs and wireless networks are discussed as well.

Chapter 3 develops an analytical model to investigate the performance of wired INs in multi-computer systems in the presence of bursty traffic with non-uniformly distributed message destinations. The model is then used to evaluate the impact of such a traffic pattern on network performance.

Chapter 4 develops an analytical model to investigate the performance of wired INs in multi-cluster systems in the presence of bursty traffic with non-uniformly distributed message destinations. By virtue of the analytical model, the effects of such a traffic pattern on the performance of INs in multi-cluster systems are investigated.

Chapter 5 presents an analytical model for wired INs in heterogeneous multi-cluster systems in the presence of bursty traffic. The analytical results provide useful insights on the trade-off between the number of clusters and the size of each cluster in heterogeneous multi-cluster systems with the aim of maximising the performance of INs.

Chapter 6 develops an analytical model to investigate the performance of integrated WLANs and WMNs in the presence of bursty traffic with non-uniformly distributed message destinations. The model is then used to obtain the maximum achievable throughput which reflects the capacity of wireless networks and to

evaluate the impact of this type of traffic on the performance of integrated wireless networks.

Chapter 7 presents an analytical performance model to investigate the QoS measures including end-to-end delay and throughput in integrated WLANs and Internet-access mesh networks in the presence of bursty traffic. The model is then used to investigate the bottleneck problem in the hybrid wireless networks.

Chapter 8 concludes the thesis and points out future directions of interest.

Chapter 2

Background Research

This chapter presents the background research and related work of wired INs and integrated wireless networks, and then describes the models characterising the nature of the realistic network traffic exhibited by real-world applications which are necessary for understanding and implementation of the models to be developed in the following chapters. Finally, the methodologies which will be used for the performance study on wired INs and integrated wireless networks are discussed.

2.1 Interconnection networks

An IN is the hardware fabrics supporting communications between individual processors in HPC systems [17, 23, 79]. A task in such systems is partitioned into smaller sub-tasks which are processed in a parallel scheme using multiple collaborating processors. The IN employs a variety of topologies that can be classified into two broad categories: *direct networks* and *indirect networks*.

In direct networks, each node has a direct connection to some of the other nodes allowing for direct communications between processors. Examples of direct INs include mesh, hypercube, torus, star graph, etc. Fig. 2.1 shows the typical instances of direct INs which have been widely employed in recent practical multi-

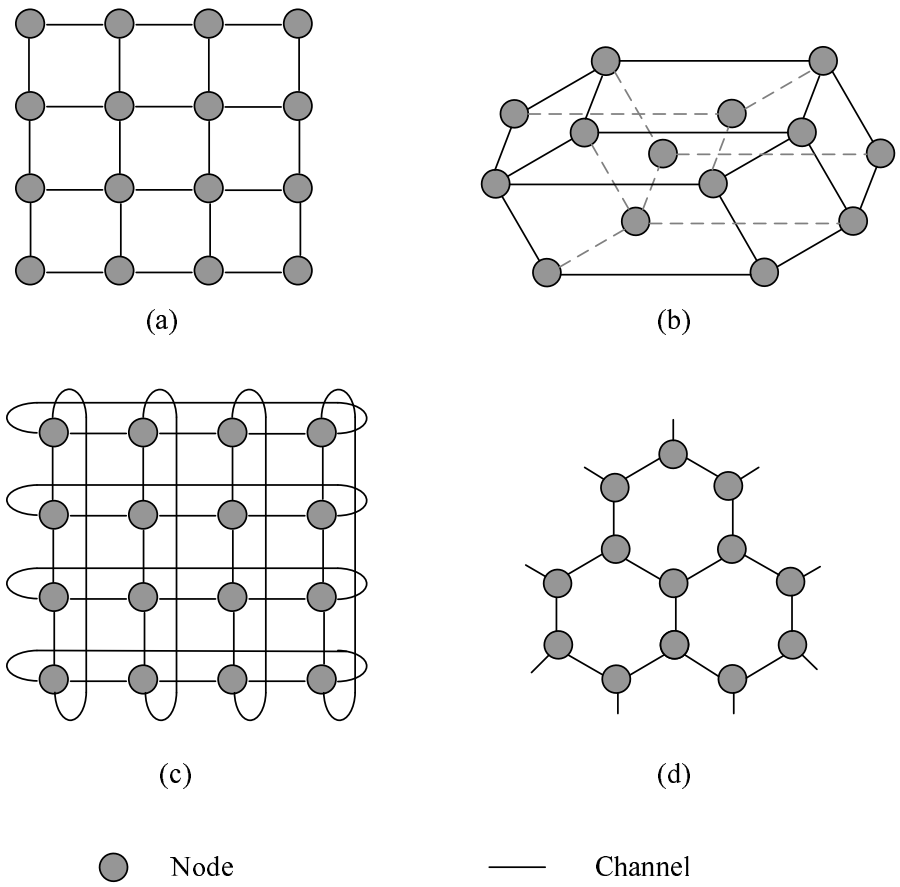

Figure 2.1: The instances of direct INs: (a) 2 dimension mesh, (b) 4 dimension hypercube, (c) 2 dimension torus, and (d) star graph

computer systems, including the Intel iPSC, Intel Delta, Intel Cavallino, Cosmic Cube, nCUBE, MIT Alewife, MIT J-Machine, MIT M-Machine, MIT Reliable Router, iWarp, Stanford FLASH, Cray T3D, Cray T3E, SGI Origin, SGI SPIDER, and Chaos Router [17].

In indirect networks, the nodes are connected to other nodes through multiple intermediate stages of switches. Examples of indirect networks include crossbar, multistage interconnection networks (MINs), fat-tree, etc. Fig. 2.2 shows the typical instances of indirect INs. Many experimental and commercial parallel machines have employed indirect INs, such as Cray X/Y-MP, DEC GIGA switch, Myrinet, NEC, Cenju-3, IBM RP3, IBM SP, Thinking Machine CM-5, Meiko CS-2, and Hitachi

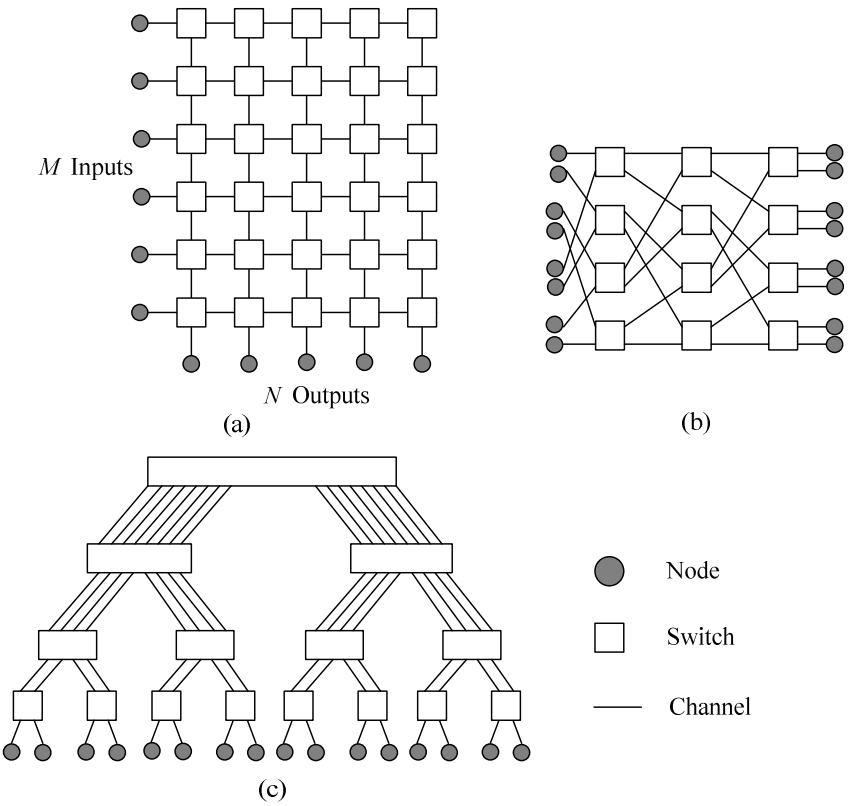

Figure 2.2: The instances of indirect INs: (a) $M \times N$ crossbar, (b) 3 stage, 2×2 Omega MINs, and (c) 16-node fat-tree

SR2201 [17]. In particular, the fat-tree topology has raised in popularity for the underlying interconnects in cluster-based systems [23, 55].

2.2 Integrated wireless networks

The recent development of wireless technologies enables access to the Internet at any place, at any time and in any form [10, 96, 117]. WMNs are a key technology to interoperate with heterogeneous wireless networks, e.g., Wi-Fi, WiMAX, WSNs, WPANs, and WLANs, in the hybrid environments and provide last-mile wireless broadband connectivity for the Internet access at a low cost [4, 9, 25, 46, 87, 97]. Mesh routers (MRs) and mesh clients (MCs) are the key components of WMNs. The WMN, in general, comprises two tiers: backhaul tier and access tier. The backhaul

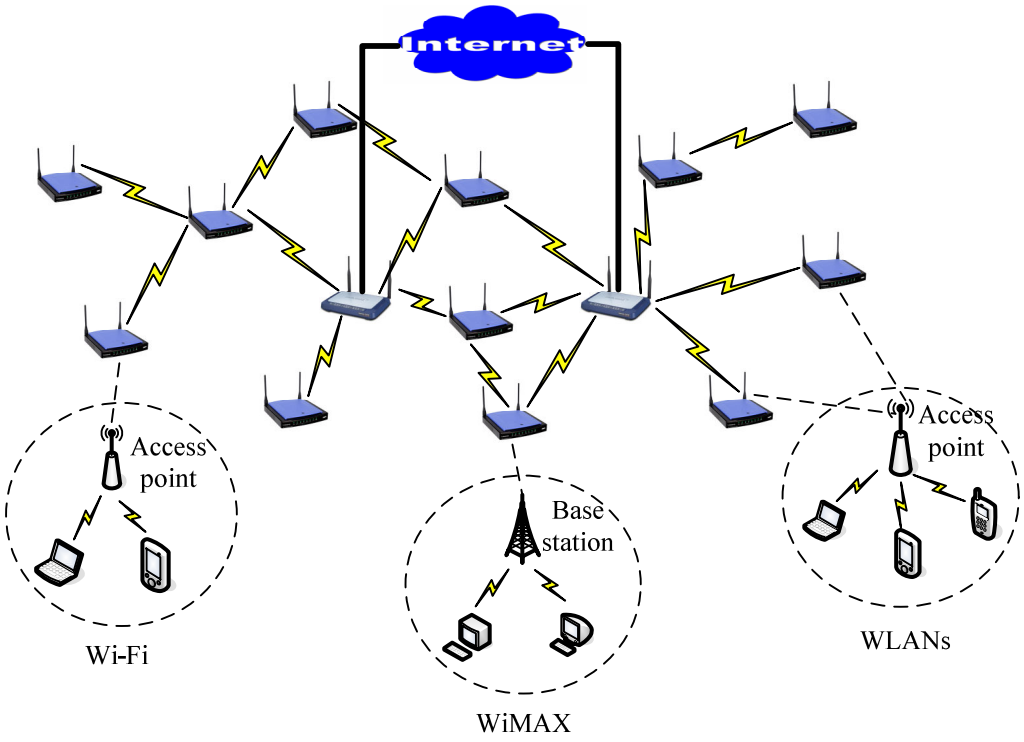

Figure 2.3: The architecture of heterogeneous wireless networks

tier contains MRs, which form a multi-hop ad hoc network and is able to be self-organised, self-configured, and self-healed to robustly interoperate in heterogeneous environments in the access tier. The mobile or static users, including pocket PCs, PDAs, laptops, desktops, etc., in the access tier are MCs. MRs have minimal mobility with sophisticated processing capabilities and contain additional routing functions, compared to the conventional wireless routers, to support mesh networking in the backhaul tier. They perform the dual tasks of both forwarding packets as well as providing network access to MCs. MRs which are facilitated with the gateway functionality can be connected to the Internet; these special MRs are called gateways.

Fig. 2.3 shows the architecture of integrated heterogeneous wireless networks with the WMN as the Web-In-the-Sky to provide the Internet access for a series of wireless networks. WMNs are becoming a major avenue for the next-generation

wireless networks since it possesses a variety of advantages including [4]: 1) a WMN can extend the coverage range of conventional single wireless networks and double the user density; 2) the modular chassis of WMNs simplifies the ability to scale; 3) the automated capabilities, such as dynamic self-organisation, quick deployment, and easy maintenance, simplify the deployment and operation; and 4) it is able to maintain the performance over multiple hops which can reduce costly connections to the optical cores.

The performance of wired or wireless networks is greatly dependent on the network topology, the switching method in wired networks or the medium access control (MAC) scheme in wireless networks, the routing strategy, and the traffic pattern exhibited by real-world applications. The specifications of network topology, switching method or MAC scheme, and the routing strategy will be presented in the following chapters as appropriate. The traffic models characterising the workloads generated by many real-world applications will be presented in the forthcoming sections of this chapter.

2.3 Traffic modelling

The traffic pattern has a significant impact on network performance. Therefore, the performance study of wired or wireless networks must incorporate accurate models of the network traffic. Since the implementation of real applications is time consuming and costly, mathematical modelling of network traffic can be used instead to evaluate the network performance at an early stage of the design process [77]. The *arrival process* and *destination distribution* of messages are the key parameters to define the network traffic patterns.

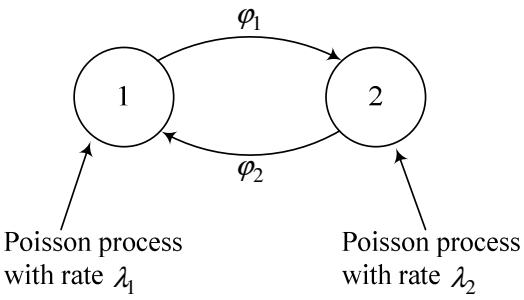

Figure 2.4: Markov model for the two-state MMPP

2.3.1 Modelling the bursty nature of message arrival processes

The Markov-modulated Poisson process (MMPP) [31] is a doubly stochastic process with an arrival rate varying according to an irreducible continuous time Markov chain and has been widely applied to model the arrival process of bursty traffic due to the following reasons:

- Numerous studies [31, 43, 107] have revealed that the MMPP has the ability of capturing the time-varying arrival rate and the important correlations among inter-arrival times;
- MMPP is closed under the splitting and superposition operations [31] and thus the MMPP can be used to model the splitting and superposition in the wired or wireless networks;
- The queueing-related results of MMPP have been widely studied [43-45, 70], which makes the solutions of modelling networks with the MMPP arrival process analytically tractable.

In this thesis, a two-state MMPP is employed to model the message arrival process of the network traffic, as shown in Fig. 2.4, which has been widely used to model the bursty nature of message arrivals [70, 82, 119]. The MMPP is

characterised by the infinitesimal generator, $\mathbf{Q}$, of the underlying Markov chain and rate matrix, $\mathbf{\Lambda}$, as

$$\mathbf{Q} = \begin{bmatrix} -\varphi_1 & \varphi_1 \\ \varphi_2 & -\varphi_2 \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{2.1}$$

where the element φ_1 is the transition rate from state 1 to 2 and φ_2 is the rate out of state 2 to 1. λ_1 and λ_2 are the mean arrival rate when the Markov chain is in state 1 and 2, respectively.

The mean arrival rate and the counting function, $N(t)$ (i.e., the number of message arrivals in time $(0,t]$), of the MMPP play an important role in the determination of the arrival process at network channels in wired or wireless networks [44].

The mean, $E[N(t)]$, variance, $Var[N(t)]$, third centralised moment, $\gamma^{(3)}(t)$, of $N(t)$ can be found as follows [44, 45]

$$E[N(t)] = \frac{\lambda_1\varphi_2 + \lambda_2\varphi_1}{\varphi_1 + \varphi_2} t \tag{2.2}$$

$$Var[N(t)] = \frac{\lambda_1\varphi_2 + \lambda_2\varphi_1}{\varphi_1 + \varphi_2} t + \frac{2(\lambda_1 - \lambda_2)^2 \varphi_1\varphi_2}{(\varphi_1 + \varphi_2)^3} t - \frac{2\varphi_1\varphi_2(\lambda_1 - \lambda_2)^2}{(\varphi_1 + \varphi_2)^4}(1 - e^{-(\varphi_1 + \varphi_2)t}) \tag{2.3}$$

$$\gamma^{(3)}(t) = E[(N(t) - E[N(t)])^3] = Z^{(3)}(1,t) - 3E[N(t)](E[N(t)] - 1)\frac{Var[N(t)]}{E[N(t)]}$$

$$- E[N(t)](E[N(t)] - 1)(E[N(t)] - 2) \tag{2.4}$$

where

$$Z^{(3)}(1,t) = \frac{6}{\varphi_1 + \varphi_2}\left\{\frac{A_{11}}{6}t^3 + \frac{A_{21}}{2}t^2 + A_{31}t + A_{12}te^{-(\varphi_1 + \varphi_2)} + A_{41}[1 - e^{-(\varphi_1 + \varphi_2)t}]\right\} \tag{2.5}$$

where the expressions for A_{ij} are given by

$$A_{11} = \frac{(\lambda_1\varphi_2 + \lambda_2\varphi_1)^3}{(\varphi_1 + \varphi_2)^2} \tag{2.6}$$

$$A_{21} = \frac{2\varphi_1\varphi_2(\lambda_1 - \lambda_2)^2(\lambda_1\varphi_2 + \lambda_2\varphi_1)}{(\varphi_1 + \varphi_2)^3} \tag{2.7}$$

$$A_{31} = \frac{\varphi_1\varphi_2(\lambda_1 - \lambda_2)^2[\lambda_1\varphi_1 + \lambda_2\varphi_2 - 2(\lambda_1\varphi_2 + \lambda_2\varphi_1)]}{(\varphi_1 + \varphi_2)^4} \tag{2.8}$$

$$A_{41} = \frac{-2\varphi_1\varphi_2(\lambda_1 - \lambda_2)^3(\lambda_1 - \lambda_2)}{(\varphi_1 + \varphi_2)^5} \tag{2.9}$$

$$A_{12} = \frac{\varphi_1\varphi_2(\lambda_1 - \lambda_2)^2(\lambda_1\varphi_1 + \lambda_2\varphi_2)}{(\varphi_1 + \varphi_2)^4} \tag{2.10}$$

The mean arrival rate, $\overline{\lambda}$, the index of dispersion for counts at time t, $\wp(t)$, and the limiting index of dispersion for counts, $\wp$, can be given by

$$\overline{\lambda} = \frac{E[N(t)]}{t} \tag{2.11}$$

$$\wp(t) = \frac{Var[N(t)]}{E[N(t)]} \tag{2.12}$$

$$\wp = \lim_{t\to\infty} \wp(t) \tag{2.13}$$

The squared coefficient of variation (SCV), C^2, of the inter-arrival time and the one-step correlation coefficient, r^1, are often used to represent the degree of burstiness of message arrivals and the correlations between inter-arrival times [70]. C^2 and r^1 of the MMPP can be given by

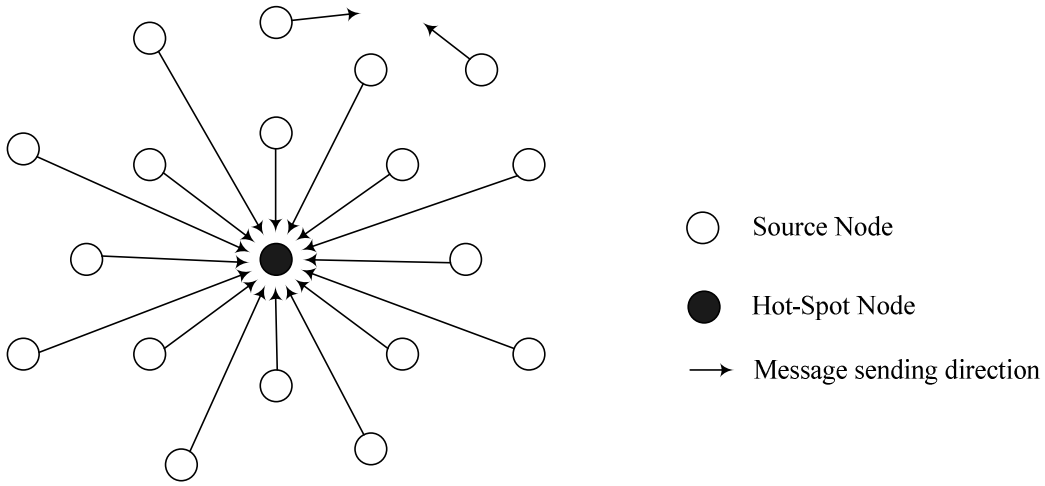

Figure 2.5: The hot-spot traffic behaviour

$$C^2 = 1 + \frac{2\varphi_1\varphi_2(\lambda_1 - \lambda_2)^2}{(\varphi_1 + \varphi_2)^2(\lambda_1\lambda_2 + \lambda_1\varphi_2 + \lambda_2\varphi_1)} \tag{2.14}$$

$$r^1 = \frac{\lambda_1\lambda_2(\lambda_1 - \lambda_2)^2\varphi_1\varphi_2}{C^2(\varphi_1 + \varphi_2)^2(\lambda_1\lambda_2 + \lambda_1\varphi_2 + \lambda_2\varphi_1)^2} \tag{2.15}$$

2.3.2 Modelling the non-uniform message destinations

The most frequently used, simplest and most elegant destination distribution is the classical uniform destination distribution where messages generated by each node are sent to other network nodes with the equivalent probability. This destination distribution has been widely used for the performance study on wired or wireless networks [5, 7, 21, 50, 51, 56, 73, 76, 91]. Based on the observations of non-uniformly distributed message destinations, extensions have been made to the uniform destination distribution. The hot-spot and locality are two typical examples of non-uniform message destination distributions and have attracted much research efforts [54, 85, 92, 102, 103, 121].

With hot-spot model [93], a large number of nodes direct a fraction of their

generated messages to the hot-spot destination node, as shown in Fig. 2.5. This node often receives a larger amount of traffic than the other network nodes, which causes the higher traffic loads on network channels located closer to the hot-spot node. The hot-spot traffic may lead to extreme network congestion resulting in serious performance degradation due to the tree saturation phenomenon [106]. Hot-spot traffic behaviour may be exhibited in many applications such as global synchronisation, cache coherency protocols, etc. It can be produced directly by certain collective communication operations.

The locality exists when the likelihood of communication to various nodes decrease with the distance, which implies that the mean inter-node distance with communication locality is smaller than that with the uniform destination distribution [23]. As a result, each message consumes fewer resources and thus reduces contention of network resources, e.g., network bandwidth. Sphere of locality [2] is widely used in modelling of the locality destination distribution, in which a processor is more likely to send messages to a small number of destinations within a so-called "sphere of locality". Destinations outside the sphere are requested less frequently.

2.4 Methodologies of performance study

The study of network performance can be achieved by three ways: test-bed, simulation, and analytical modelling [17, 23]. The test-bed and simulation based approach may be time-consuming and costly, especially for large-scale systems. In contrast, analytical modelling can capture the essential features of the system, gain significant insights into the system performance, and offer a cost-effective and versatile tool that can be used to investigate the system performance under different design alternatives and various working conditions.

In particular, the analytical models can provide quantitative relations between input parameters and performance metrics to have a thorough investigation of the system performance over a complete parametric range [77]. Realising such detailed investigations through simulation may take months or years, depending on the machine deployed, while an efficient analytical model can provide a great reduction in the time required for such investigations. Most analytical models that are proposed to investigate the network performance use the results from the queueing theory, providing a relatively straightforward derivation for the expressions of performance metrics.

2.5 Related work

2.5.1 Performance studies of interconnection networks in multi-computer systems

Most related analytical models [5, 21, 73, 91] on the performance of INs in multi-computer systems reported in the literature are based on the assumptions that the message arrivals follow the non-bursty Poisson process and the message destinations are uniformly distributed over all network nodes. The authors [72] proposed an analytical model for the deterministic wormhole-switched torus networks. In deterministic routing, messages with the same source and destination addresses always take the same network route. Consequently, they cannot use alternative paths that may avoid blocking. Fully adaptive routing has been proposed to overcome this limitation by enabling messages to explore all available paths in the network. To capture the performance benefits of adaptive routing in INs, Ould-Khaoua [91]

developed a performance model for wormhole-switched uni-directional torus networks with Duato's fully adaptive routing. In this model, the message arrivals are modelled by Poisson process and message destinations are uniformly distributed over the network nodes. Due to the advances of bi-directional technology in real systems [17], the authors [73] proposed an analytical model for adaptive wormhole routing in bi-directional torus networks. This model also considers the non-bursty message arrivals with uniformly distributed destinations.

To capture the characteristics of realistic network traffic, most researches have to use simulation to evaluate the performance of INs under bursty traffic loads and/or non-uniformly distributed destinations. For instance, Pombortsis and Halatsis [94] studied the hot-spot phenomenon in shared memory parallel computers. Kotapati and Dandamudi [63] employed the simulation to study the buffer management issues in wormhole-switched torus networks. This model considered the effects of hot-spot traffic on network performance and proposed a hybrid buffer organisation scheme that combines both centralised and distributed methods to accommodate this type of traffic pattern. Orduna and Duato [90] studied a switching method that combines both wormhole switching and circuit switching in the parallel computing systems under bursty traffic. Shin and Pinkston [109] used the simulation experiments to investigate the routing algorithms for INs in multi-computer systems under bursty traffic with non-uniformly distributed message destinations.

Several recent studies have explored to develop analytical models to handle the bursty traffic or the non-uniformly distributed message destinations, separately. For instance, performance models [82, 122] were developed to predict the message latency of INs in multi-computer systems taking the bursty nature of message arrivals into consideration. However, these models consider the uniform message

destination distributions. On the other hand, the authors [92, 102, 103] proposed performance models for INs in multi-computer systems considering the hot-spot destinations. These models make the assumption that the message arrivals follow the non-bursty Poisson process without capturing the impact of bursty traffic on network performance. However, there has not been any analytical model reported in the current literature to handle the bursty traffic and non-uniformly distributed message destinations simultaneously which can capture the bursty nature of realistic network traffic in the both temporal domain and spatial domain.

2.5.2 Performance studies of interconnection networks in multi-cluster systems

The performance studies of INs in multi-cluster systems have been reported in the literature. Abawajy [1] proposed an adaptive space-sharing scheduling policy for heterogeneous commodity-based cluster systems to minimise the impacts of heterogeneity and processor load variation on network performance of cluster systems. Lin, Chung, and Huang [69] proposed a routing algorithm in fat-tree based INs and developed a software simulator to evaluate the performance benefits of the proposed algorithm in the network. The authors [36] proposed a deterministic routing algorithm for fat-tree based INs and compared it with the adaptive routing under different traffic conditions. The simulation results showed that the deterministic routing can achieve a similar, and in some scenarios higher, level of performance than adaptive routing in fat-trees.

Analytical models for the performance study of INs in multi-cluster systems are rarely found in the literature. Very recently, the authors [56] proposed an analytical model for the performance study of INs in multi-cluster systems. However,

the models assume that the network traffic follows the non-bursty Poisson process and the message destinations are uniformly distributed over the network nodes. A more comprehensive model [54] was developed to investigate the network performance in multi-cluster systems taking into consideration the communication locality which is a typical example of non-uniform destination distributions. However, this model makes the simplified assumption that message arrivals follow the non-bursty Poisson process. The authors [84] recently proposed an analytical model to capture the hot-spot phenomenon in fat-tree based INs in multi-cluster systems with the message arrivals modelled by non-bursty Poisson process.

In addition, most existing researches are based on homogeneous cluster systems and the evaluations are confined to a single cluster [22, 59] with the exception of [14]. The authors in [14] investigated the performance of a cluster-based system taking the heterogeneity of processors into consideration. The analytical models on the performance of heterogeneous multi-cluster systems are generally rare in the literature, in particular when the network is subject to the bursty traffic. To the best of author's knowledge, there is only one analytical model [24] to evaluate the network performance of multi-cluster systems taking the bursty traffic into account. However, in this model, each cluster is modelled by an infinite-buffer queueing system without considering the specific structure and network components in the cluster, e.g., network topology, message switching and routing scheme, etc.

2.5.3 Performance studies of integrated wireless networks

The performance modelling and evaluation of WLANs have received numerous attentions. Bianchi [7] presented an analytical model for IEEE 802.11 WLANs based on distributed coordination function (DCF) and derived the network throughput with

the assumption of homogeneous stations and ideal channels under saturated traffic conditions. The proposed analysis was able to apply to both packet transmission schemes (i.e., basic access and RTS/CTS access mechanisms) employed by DCF. Since the typical network conditions are non-saturated and heterogeneous, Malone, Duffy, and Leith [76] presented an extension work of Bianchi's model to the heterogeneous stations with non-saturated traffic conditions. Although the IEEE 802.11 has been widely used to deploy WLANs, it lacks of QoS support in the protocol. To support QoS in WLANs, the IEEE 802.11e has been standardised and received significant attentions recently. A performance model for WLANs based on IEEE 802.11e enhanced distributed channel access (EDCA) was presented in [51]. The model took into account the QoS support of EDCA and derived the packet delay and network throughput under the saturated traffic conditions. Hu, Min, and Woodward [50] proposed an analytical model for evaluating the throughput, packet delay, and loss probability of IEEE 802.11e-based WLANs under the non-saturated traffic conditions. However, all these studies are based on the simplified assumptions that the traffic generated by stations follows the non-bursty Poisson process. Very recently, the authors [80] proposed an analytical model for WLANs based on IEEE 802.11e EDCA taking the bursty nature of realistic traffic patterns into account and investigated the performance benefits of transmission opportunity mechanism of EDCA under this type of traffic pattern.

Analytical models for WMNs have also been widely reported in the literature. Cao, Ma, and Zhang [11] developed a model to analyse distributed mesh mode scheduler used in the IEEE 802.16 protocol. Bisnik and Abouzeid [8] developed an analytical model for end-to-end delay in WMNs using a random access MAC protocol like IEEE 802.11 DCF. The model was then applied to obtain the maximum

achieve throughput in WMNs. Liu and Liao [71] provided a framework to study the location-dependent throughput and delay of WMNs by considering the unidirectional traffic under the saturated condition. Le and Hossain [66] proposed a tandem queueing model for the performance analysis and engineering in WMNs. They also investigated the problem of QoS routing in multi-hop wireless networks by means of the developed analytical model. Huang, Wang, and Chang [52] investigated the trade-off among QoS, capacity, and coverage in a ring-based WMN. They further developed a PHY/MAC cross-layer analytical model to evaluate the delay, jitter, and throughput of the proposed WMNs. An analytical model based on the single node decomposition methodology was proposed in [128] to investigate the performance of WMNs. This model took into account the interference due to the neighbour transmissions. However, these models are mainly developed under the assumptions that the traffic generated by the stations follows the non-bursty Poisson process. Several studies resorted to simulations or analytical modelling to evaluate the performance of WMNs when the network is subject to bursty traffic. For instance, the authors [113] recently presented simulation results for the investigation of end-to-end delay and throughput in WMNs in the presence of bursty traffic. Fallahi, Hossain, and Alfa [26] proposed a queueing analytical framework to study the trade-off between the energy saving and the QoS at the relay node in a energy-limited mesh networks under bursty traffic.

Most of analytical models on the performance study are carried out in a single wireless network (e.g., WLANs) or WMNs only. Recently, the network selection problem [88], radio resource management [47], admission control [28], etc., in heterogeneous wireless networks have been addressed and reported in the literature. Niyato and Hossain [87] presented a bandwidth management and admission control

framework for the integration of WLANs with multi-hop infrastructure mesh networks based on the Game-Theoretic approach. However, these models cannot be used to investigate the QoS performance measures in the integrated wireless networks. The main problem in the current literature is the lack of researches regarding the delay and throughput analysis in such a hybrid wireless environment. Moreover, there is hardly any model reported to handle the bursty traffic with non-uniformly distributed message destinations in integrated wireless networks.

Chapter 3

A Performance Model for Interconnection Networks in Multi-computer Systems under Bursty Traffic with Hot-spot Destinations

A number of measurement studies [19, 93, 108, 111] have convincingly demonstrated that the traffic generated by many real-world applications in INs of multi-computer systems exhibits bursty nature and the message destinations are non-uniformly distributed over the network nodes. Such traffic patterns have significantly different theoretical properties from the traditional Poisson arrival process with uniformly distributed message destinations [62, 103]. Many analytical models for INs in multi-computer systems have been widely reported [5, 16, 21, 41, 59, 60, 73, 91]. Most of these models are based on the simplified assumptions that the message arrivals follow non-bursty Poisson process and the message destinations are uniformly distributed over all network nodes. Several recent studies have explored to develop analytical models to handle the bursty message arrivals [64, 82, 122] or the non-uniform message destination distributions [92, 102, 103], separately.

However, there has not been any study reported in the current literature to handle the traffic burstiness in the both temporal domain (i.e., bursty message arrival

process) and spatial domain (i.e., non-uniform message destination distributions) simultaneously. With the aim of capturing the characteristics of realistic traffic patterns and obtaining a comprehensive understanding of the performance behaviour of INs in multi-computer systems, this chapter presents an analytical model to investigate the performance of INs in multi-computer systems in the presence of bursty traffic with hot-spot destinations. The traffic characteristics on network channels are determined in a hop-by-hop scheme with respect to the distance to the hot-spot node, due to the non-uniform distribution of traffic on network channels. Extensive simulation experiments are conduced to validate the accuracy of the model. The model is then applied to investigate the impact of bursty traffic with hot-spot destinations on the performance of INs in multi-computer systems.

The rest of this chapter is organised as follows. Section 3.1 shows useful preliminaries including the topology of INs, the switching method and routing strategy for message flow control, and the hot-spot destination distributions. Section 3.2 presents the assumptions and notations adopted in this chapter. Section 3.3 derives the analytical performance model to investigate the performance of INs in multi-computer systems. Extensive simulation experiments are used to validate the accuracy of the analytical model in Section 3.4. Section 3.5 carries out performance analysis. Finally, Section 3.6 concludes this study.

3.1 Preliminaries

3.1.1 Interconnection networks in multi-computer systems

Current multi-computer systems have widely employed torus as their underlying IN for low-latency and high-bandwidth inter-processor communications due to its

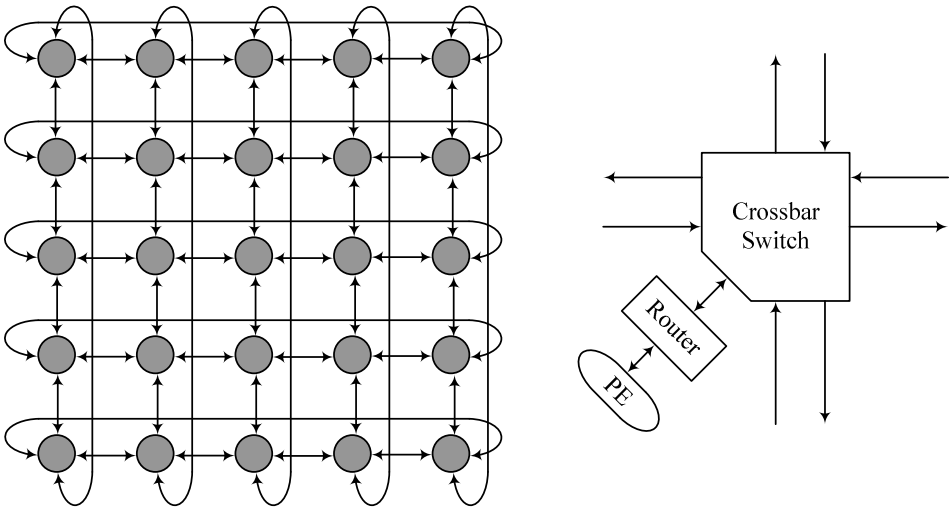

Figure 3.1: (a) A 2-dimensional torus network and (b) the node structure

attractive properties including symmetry and regularity structure, constant node degree, modularity, and the ability to exploit communication locality found in many parallel applications [17]. A torus can be expanded simply by adding nodes and channels without any change to the existing node structure.

A 2-dimensional torus consists of $N = k^2$ nodes in a regular grid structure with k nodes in each dimension [23]. Each node can be identified by a 2-digit radix k address (x, y), $0 \le x, y \le k-1$, where x represents the row number and y indicates the column number of the node. A node with the address (x, y) is interconnected to nodes $(x \pm 1, y)$ modulo k and nodes $(x, y \pm 1)$ modulo k. Fig. 3.1(a) depicts a 2-dimensional torus with 5 nodes in each dimension. In torus, wraparound channels are used to connect each node on the edge to the corresponding node on the opposite edge.

Each node in torus consists of a processing element (PE) and a router, as shown in Fig. 3.1(b). The router has 5 inputs and 5 output channels. Each node is connected to its 4 neighboring nodes through 4 inputs and 4 output channels. The

remaining channels are used by the PE to inject/eject messages to/from the network, respectively. The input and output virtual channels are connected by a $5V$ -way crossbar switch, with V being the number of virtual channels which shares the bandwidth of the physical channel with other virtual channels in a time-multiplexed fashion [16].

3.1.2 Switching methods

The switching method determines the way messages visit intermediate nodes and deals with the allocation of channel and buffer resources to a message as it travels over the network. Packet switching and wormhole switching are two of the most important flow control mechanisms used in commercial routers [23]. With packet switching, each node waits until a packet has been completely received and then forwards the packet to the next node. The main disadvantage of this switching mechanism is its high latency. In contrast, wormhole switching overcomes the latency penalty of packet switching by dividing a message into a sequence of fixed-size units, called flits, each of a few bytes for transmission and flow control. The channel and buffer need to be allocated to flits only rather than the whole packet. This mechanism has been popular and will be adopted in this study due to its low buffering requirements and, more importantly, it makes end-to-end latency insensitive to the message distance in the absence of blocking. Moreover, wormhole switching has been an attractive design alternative not only in contemporary network-based systems but also in the new network-on-chip (NoC) and system-on-chip (SoC) architectures [77].

In this flow control mechanism, the header flit governs the path through the network and the remaining data flits follow it in a pipelined fashion. If the outgoing

channel is busy, the message is blocked in situ until the outgoing channel becomes available. A blocked message may occupy a fraction of network resources along its path. Consequently, deadlocks are possible unless a deadlock-free routing strategy is employed. The use of virtual channels on a physical channel can facilitate the development of deadlock-free routing and also improve the bandwidth utilisation of physical channels [16].

3.1.3 Routing strategies

The routing algorithm specifies how a message selects its network path to cross from the source to the destination. Adaptive routing use information about network traffic and/or channel status to allow messages to be no longer restricted to travel along the same path but to be directed over all available channels from the source to the destination in order to avoid congested or faulty regions in the network. Duato's fully adaptive routing algorithm [23] divides the virtual channels into two classes. At each routing step, a message can adaptively visit any available virtual channel from the first class. If all the virtual channels belonging to the first class are busy, it crosses a virtual channel from the second class using the deterministic routing. The virtual channels of the second class define a complete virtual deadlock-free sub-network built from the virtual channels of the first class.

3.1.4 Non-uniform message destination distributions

Hot-spot is able to capture the characteristics of non-uniformly distributed message destinations where a number of nodes direct a fraction of their messages to the hot-spot node [93]. This node often receives a larger amount of traffic than the other network nodes, which causes the higher traffic loads on network channels located

closer to the hot-spot node. The hot-spot traffic may lead to paralysis of the hot-spot node and even the whole network [124]. Hot-spot traffic has attracted significant research efforts over the past few years due to the strong evidence of its existence and the great efforts on network performance [42, 74, 92, 103]. For example, messages routed through an IN with the same destination address (i.e., the network coordinator) may result in contention [108]. Moreover, global synchronization, where each node sends a synchronisation message to a specific node which coordinates the synchronisation activity of the system, is a typical situation that can produce hot-spots [93, 98].

The hot-spot traffic model proposed in [93] is employed to generate non-uniform distribution of message destinations in this chapter. Specifically, each message has the probability, δ, to be directed to the hot-spot node (i.e., hot-spot messages), and the probability, $(1-\delta)$, of being evenly directed to all network nodes (i.e., regular messages).

3.2 Assumptions and notations

The model is based on the following assumptions [5, 16, 21, 41, 59, 60, 73, 81, 82, 91, 92, 102, 103, 121, 122], where Assumptions (a) and (b) are distinguishing this study from the existing ones to model the bursty traffic with non-uniformly distributed destinations. A brief summary of key notations used in the derivation of the model is listed in Table 3.1.

a) The arrivals of messages generated by the source node are bursty and follow a two-state MMPP_s (the subscript s denoting the traffic generated by the source node). The MMPP_s can be characterised by the infinitesimal generator, $\mathbf{Q}_s$, of

the underlying Markov chain and rate matrix, $\mathbf{\Lambda}_s$. $\mathbf{Q}_s$ and $\mathbf{\Lambda}_s$ can be given by

$$\mathbf{Q}_s = \begin{bmatrix} -\varphi_{s1} & \varphi_{s1} \\ \varphi_{s2} & -\varphi_{s2} \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda}_s = \begin{bmatrix} \lambda_{s1} & 0 \\ 0 & \lambda_{s2} \end{bmatrix} \tag{3.1}$$

b) The destinations of messages are non-uniformly distributed over the network nodes and follow the hot-spot distribution depicted in Chapter 3.1.4;

c) The message length is ω flits. Each flit requires one-cycle transmission time to cross a physical channel;

d) The physical channel is partitioned into V $(V > 2)$ virtual channels. In Duato's routing algorithm [23], the first class contains $(V - 2)$ virtual channels, and the second class contains 2 virtual channels.

Table 3.1: Key notations used in the derivation of the model in Chapter 3

k, N	number of nodes per dimension and the network size
V	total number of virtual channels per physical channel
δ	hot-spot fraction, i.e., the probability that each generated message is sent to the hot-spot node
ω	message length in flits
$\overline{T}$	mean network latency
$\overline{W}_s$	mean waiting time at the source node
$\overline{V}$	mean degree of virtual channel multiplexing at a given physical channel
$\overline{Latency}$	mean message latency
$\overline{\eta}_r$, $\overline{d}_r$	average number of hops that a regular message visits along each of the 2 dimensions and across the network
t_r, t_{h_j}	network latency for a regular message and a j-hop hot-spot message

$F_{t_r}^*(s)$, $F_{t_{h_j}}^*(s)$	Laplace-Stieltjes transforms of t_r and t_{h_j}
t_h	mean network latency for hot-spot messages
b_{r_i}	blocking time experienced by a regular message at its *i*-th hop channel $(1 \le i \le \bar{d}_r)$
$b_{h_{i,j}}$	blocking time for a *j*-hop hot-spot message at its *i*-th hop channel $(1 \le i \le j)$
$F_{b_r}^*(s)$, $F_{b_{h_j}}^*(s)$	Laplace-Stieltjes transforms of blocking time for a regular message and a *j*-hop hot-spot message
P_j	the probability that a message needs to make j hops to reach its destination
N_j	the number of nodes located j hops away from a given node
ξ_j	the probability that a network channel is located at j hops away from the hot-spot node
C_j	the number of channels located j hops away from the hot-spot node
$Pb_{i,j}$	the probability that a message is blocked at its *i*-th hop channel located j hops away from the hot-spot node
P_{a_j}	the probability that all adaptive virtual channels at a physical channel located j hops away from the hot-spot node are busy
$P_{a\&d_j}$	the probability that all adaptive and deterministic virtual channels of the physical channel are busy
$f_{b_r}(x)$, $F_{b_r}^*(s)$	the probability density function of the blocking time for a regular message and its Laplace-Stieltjes transform
$\mathbf{Q}_s$, $\mathbf{\Lambda}_s$	parameter matrices of MMPP_s to model the traffic generated by the source node
$\mathbf{Q}_{cr}$, $\mathbf{\Lambda}_{cr}$	parameter matrices of MMPP_{cr} to model the regular traffic on network channels
$\mathbf{Q}_{ch_j}$, $\mathbf{\Lambda}_{ch_j}$	parameter matrices of MMPP_{ch_j} to model the hot-spot traffic on network channels located j hops away from the hot-spot node

$\mathbf{Q}_{c_j}$, $\mathbf{\Lambda}_{c_j}$	parameter matrices of the MMPP_{c_j} to model the superposed traffic on network channels located j hops away from the hot-spot node
$\mathbf{Q}_v$, $\mathbf{\Lambda}_v$	parameter matrices of the MMPP_v to model the traffic arriving at an injection virtual channel
t_{c_j}	the service time at network channels located j hops away from the hot-spot node
$\boldsymbol{\pi}_{c_j}$	the steady state vector of MMPP_{c_j}
$P_{v,j}$	the probability that v virtual channels at a given physical channel located j hops away from the hot-spot node are busy
t_{c_j} , $t_{c_j}^{(2)}$	the first two moments of the service time on network channels located j hops away from the hot-spot node
$F_{t_{c_j}}^{*}(s)$	Laplace-Stieltjes transforms of t_{c_j}
W_{s_j}	the waiting time at the source node located j hops away from the hot-spot node

3.3 The analytical model

The message latency consists of: a) the network latency, $\overline{T}$, that is the time for a message to cross the network, and b) the delay experienced by a message at the source node, $\overline{W_s}$. In order to model the effect of virtual channel multiplexing, the latency has to be scaled by a factor, $\overline{V}$, representing average degree of virtual channel multiplexing that takes place at a given physical channel [91]. Thus, the message latency can be written as

$$\overline{Latency} = (\overline{T} + \overline{W}_s)\overline{V} \tag{3.2}$$

Taking both the regular and hot-spot messages with their corresponding probabilities into account, the network latency can be given by

$$\overline{T} = (1-\delta)t_r + \delta\, t_h \tag{3.3}$$

3.3.1 Network latency for regular messages

The regular and hot-spot messages experience different network latencies due to the non-uniform traffic loads and diverse waiting time over different network channels with the locations respect to the hot-spot node.

Since adaptive routing distributes the regular messages evenly across the network channels, the average number of hops that a regular message visits in a given dimension, $\overline{\eta}_r$ (the subscript r representing the regular messages), and across the network, $\overline{d}_r$, can be given by [17]

$$\overline{\eta}_r = \begin{cases} \dfrac{k}{4} & k \text{ is even} \\ \dfrac{1}{4}\left(k - \dfrac{1}{k}\right) & k \text{ is odd} \end{cases} \quad \text{and} \quad \overline{d}_r = 2\overline{\eta}_r \tag{3.4}$$

Due to the symmetry of the torus topology and the use of adaptive routing, the regular message is destined evenly across the network. Therefore, the network latency experienced by a regular message, t_r, can be determined by the actual message transmission time and the delay due to blocking in the network. t_r can be written as

$$t_r = \omega + \overline{d}_r + \sum_{i=1}^{\overline{d}_r} b_{r_i} \tag{3.5}$$

where b_{r_i} is the blocking time experienced by a regular message at its i-th hop channel ($1 \le i \le \overline{d}_r$). The blocking time, b_{r_i}, can be determined by the probability, $Pb_{i,j}$, that a message is blocked at its i-th hop channel located at j hops away from

the hot-spot node and the waiting time, W_{c_j}, for a regular message to acquire a deterministic virtual channel at the lowest dimension when blocking occurs.

3.3.2 Network latency for hot-spot messages

In the presence of the hot-spot traffic, the channels located closer to the hot-spot node receive higher traffic loads than those farther away. Based on the determination of the network latency for the regular message, the network latency, t_{h_j}, experienced by a *j*-hop hot-spot message can be given by

$$t_{h_j} = \omega + j + \sum_{i=1}^{j} b_{h_{i,j}} \tag{3.6}$$

where $b_{h_{i,j}}$ is the blocking time for a *j*-hop hot-spot message at its *i*-th hop channel $(1 \le i \le j)$.

The probability, P_j, that a message need make j hops to reach its destination can be determined by

$$P_j = \frac{N_j}{N-1} \tag{3.7}$$

where N_j is the number of nodes located at j hops away from a given node in torus and can be given by [102]

$$N_j = \begin{cases} 1 & j = 0 \\ 4j & j < \lfloor k/2 \rfloor \\ 2(k-1) & j = \lfloor k/2 \rfloor \\ 4(k-j) & j > \lfloor k/2 \rfloor \end{cases} \tag{3.8}$$

Averaging over all the possible hops j made by the hot-spot messages yields the network latency, t_h, for the hot-spot messages as

$$t_h = \sum_{j=1}^{d_{max}} P_j t_{h_j} \tag{3.9}$$

3.3.3 The blocking time on network channels for regular messages and hot-spot messages

When the typical regular message has reached its *i*-th hop channel along its network path, this channel can be from 1 to d_{max} hops away from the hot-spot node in torus, where

$$d_{max} = 2\lfloor k/2 \rfloor \tag{3.10}$$

Let ξ_j denote the probability that a network channel is located at j $(1 \le j \le d_{max})$ hops away from the hot-spot node. ξ_j can be given by

$$\xi_j = \frac{C_j}{4N} \tag{3.11}$$

where C_j is the number of channels located at j hops away from the hot-spot node and can be found in [102]. Therefore, the blocking time, b_{r_i}, experienced by a regular message can be written as

$$b_{r_i} = \sum_{j=1}^{d_{max}} \xi_j P b_{i,j} W_{c_j} \tag{3.12}$$

A regular message is blocked at a network channel when all the adaptive virtual channels of the remaining dimensions to be visited, and also the deterministic virtual channels of the lowest dimension still to be crossed are busy [23]. Let P_{a_j} denote the probability that all adaptive virtual channels at a physical channel located

at j hops away from the hot-spot node are busy and $P_{a\&d_j}$ represent the probability that all adaptive and deterministic virtual channels of the physical channel are busy. In the two-dimensional torus, the probability that a message has both dimensions to cross is $(1-1/\overline{\eta}_r)^2$, where $1/\overline{\eta}_r$ is the probability that a regular message terminates at the current dimension. Thus, the blocking probability for such a message can be given by $(1-1/\overline{\eta}_r)^2 P_{a\&d_j} P_{a_j}$. The probability that a message has finished crossing one dimension and needs to visit the other one is $2(1-1/\overline{\eta}_r)(1/\overline{\eta}_r)$. The blocking probability for such a message can be given by $2(1-1/\overline{\eta}_r)(1/\overline{\eta}_r)P_{a\&d_j}$. Therefore, the blocking probability for a regular message, $Pb_{i,j}$, can be expressed as [103]

$$Pb_{i,j} = \left(1-\frac{1}{\overline{\eta}_r}\right)^2 P_{a\&d_j} P_{a_j} + 2\left(1-\frac{1}{\overline{\eta}_r}\right)\frac{1}{\overline{\eta}_r} P_{a\&d_j} \tag{3.13}$$

Let $P_{v,j}$, $(0 \le v \le V)$ and $(1 \le j \le d_{max})$, denote the probability that v virtual channels at a given physical channel located at j hops away from the hot-spot node are busy. The probabilities P_{a_j} and $P_{a\&d_j}$ can be given by [91]

$$P_{a_j} = P_{V,j} + \frac{2P_{(V-1),j}}{\binom{V}{V-1}} + \frac{P_{(V-2),j}}{\binom{V}{V-2}} \tag{3.14}$$

$$P_{a\&d_j} = P_{V,j} + \frac{2P_{(V-1),j}}{\binom{V}{V-1}} \tag{3.15}$$

Considering a *j*-hop hot-spot message reaching its *i*-th hop channel which is $(j-i+1)$ hops away from the hot-spot node. Let $Pb_{i,j-i+1}$ represent the probability that the *j*-hop hot-spot message is blocked at its *i*-th hop channel and $W_{c_{j-i+1}}$ denote

the waiting time experienced by a *j*-hop hot-spot message to acquire a deterministic virtual channel at the lowest dimension when blocking occurs. Therefore, the blocking time, $b_{h_{i,j}}$, can be given by

$$b_{h_{i,j}} = Pb_{i,j-i+1} W_{c\,j-i+1} \tag{3.16}$$

where the blocking probability of the hot-spot message, $Pb_{i,j-i+1}$, can be obtained by simply substituting j by $(j-i+1)$ and $\overline{\eta}_r$ by $j/2$ in the Eqs. (3.13)-(3.15).

3.3.4 Traffic characteristics on network channels

To obtain $P_{v,j}$, $(0 \le v \le V)$ and $(1 \le j \le d_{max})$, denoting the probability that v virtual channels at a given physical channel located j hops away from the hot-spot node are busy, used in Eqs. (3.14) and (3.15), the traffic characteristics on network channels need to be identified firstly. Let us refer to the loads formed by the regular messages and hot-spot messages as the regular traffic and hot-spot traffic, respectively. The loads at network channels located at j $(1 \le j \le d_{max})$ hops away from the hot-spot node are composed of both the regular traffic and hot-spot traffic.

Since adaptive routing distributes regular messages evenly on network channels, the arrival process of regular traffic at network channels exhibits similar statistical behaviours. Since each regular message visits, on average, $\overline{d}_r$ channels to reach its destination and each node has 4 output channels, the amount of traffic arriving at each network channel is, on average, f_r times of the traffic generated by a source node. f_r can be given by

$$f_r = \frac{N\overline{d}_r(1-\delta)}{4N} = \frac{\overline{d}_r}{4}(1-\delta) \tag{3.17}$$

Generally, f_r is not an integer value due to its determination from the network size, the properties of traffic generated by the source node, as well as the hot-spot fraction. Let Z_r and F_r denote the integral and fractional parts of f_r. Given that the superposition of multiple MMPPs is again an MMPP [31], let MMPP_{cr} model the regular traffic arriving at a given network channel. MMPP_{cr} can be determined by the superposition of Z_r traffic flows modelled by MMPP_s and one traffic flow modelled by MMPP_F, where MMPP_F represents the resulting traffic flow from the splitting of MMPP_s with the splitting probability F_r. According to the principle of splitting an MMPP [31], the corresponding infinitesimal generator, $\mathbf{Q}_F$, and rate matrix, $\mathbf{\Lambda}_F$, of MMPP_F can be given by

$$\mathbf{Q}_F = \mathbf{Q}_s = \begin{bmatrix} -\varphi_{F1} & \varphi_{F1} \\ \varphi_{F2} & -\varphi_{F2} \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda}_F = F_r \mathbf{\Lambda}_s = \begin{bmatrix} \lambda_{F1} & 0 \\ 0 & \lambda_{F2} \end{bmatrix} \tag{3.18}$$

By virtue of the method presented in [45], MMPP_{cr} is chosen to match the following four statistical characteristics of the regular traffic on network channels:

1) mean arrival rate;

2) index of dispersion for counts at time t;

3) limiting index of dispersion;

4) third centralised moment at time t.

According to Eqs. (2.2)-(2.13), the above four statistical characteristics, λ_F, $\wp_F(t)$, $\wp_F$, and $\gamma_F^{(3)}(t)$ of MMPP_F can be determined. The four statistical characteristics of the regular traffic on network channels generated by Z_r MMPP_s and one MMPP_F can be further obtained as follows [45]

$$\lambda_{cr} = Z_r \lambda_s + \lambda_F \tag{3.19}$$

$$\wp_{cr}(t) = Z_r \wp_s(t) \frac{\lambda_s}{\lambda_{cr}} + \wp_F(t) \frac{\lambda_F}{\lambda_{cr}} \tag{3.20}$$

$$\wp_{cr} = Z_r \wp_s \frac{\lambda_s}{\lambda_{cr}} + \wp_F \frac{\lambda_F}{\lambda_{cr}} \tag{3.21}$$

$$\gamma_{cr}^{(3)}(t) = Z_r \gamma_s^{(3)}(t) + \gamma_F^{(3)}(t) \tag{3.22}$$

where

$$t = \frac{1}{2(N+1)} \left\{ N \left(\frac{1}{\varphi_{s1}} + \frac{1}{\varphi_{s2}} \right) + \left(\frac{1}{\varphi_{F1}} + \frac{1}{\varphi_{F2}} \right) \right\} \tag{3.23}$$

The infinitesimal generator, $\mathbf{Q}_{cr}$, and rate matrix, $\mathbf{\Lambda}_{cr}$, of MMPP_{cr} can be determined by solving the following system of equations (i.e., Eqs. (3.24)-(3.27))

$$\frac{\lambda_{s1}\varphi_{s2} + \lambda_{s2}\varphi_{s1}}{\varphi_{s1} + \varphi_{s2}} = \lambda_{cr} \tag{3.24}$$

$$\frac{2(\lambda_{s1} - \lambda_{s2})^2 \varphi_{s1}\varphi_{s2}}{(\varphi_{s1} + \varphi_{s2})^2 (\lambda_{s1}\varphi_{s2} + \lambda_{s2}\varphi_{s1})} = \wp_{cr} - 1 \tag{3.25}$$

$$\frac{1 - e^{(\varphi_{s1} + \varphi_{s2})t}}{(\varphi_{s1} + \varphi_{s2})t} = \frac{\wp_{cr} - \wp_{cr}(t)}{\wp_{cr} - 1} \tag{3.26}$$

$$Z_s^{(3)}(1,t) = \lambda_{cr}^3 t^3 + 3\lambda_{cr}^2 (\wp_{cr} - 1) t^2 + \frac{3\lambda_{cr}(\wp_{cr} - 1)}{\tau} \left[\frac{(\lambda_{s1} - \lambda_{s2})(\varphi_{s1} - \varphi_{s2})}{\tau} - \lambda_{cr} \right] t$$

$$+ \frac{3\lambda_{cr}}{\tau^2} (\wp_{cr} - 1)[(\lambda_{s1} - \lambda_{s2})(\varphi_{s1} - \varphi_{s2}) + \lambda_{cr}\tau] t e^{-\tau t}$$

$$- \frac{6\lambda_{cr}}{\tau^3} (\wp_{cr} - 1)(\lambda_{s1} - \lambda_{s2})(\varphi_{s1} - \varphi_{s2})(1 - e^{-\tau t}) \tag{3.27}$$

where $Z_s^{(3)}(1,t)$ can be obtained based on Eq. (2.5) and τ can be determined by successive substitution in the following equation

$$\tau = \frac{1}{t}\frac{\wp_{cr} - 1}{\wp_{cr} - \wp_{cr}(t)}(1 - e^{-\tau t}) \tag{3.28}$$

Eq. (3.27) can be re-written in the following form

$$(\lambda_{s1} - \lambda_{s2})(\varphi_{s1} - \varphi_{s2}) = \psi \tag{3.29}$$

where ψ is known. Using the method provided in [45], the infinitesimal generator, $\mathbf{Q}_{cr}$, and rate matrix, $\mathbf{\Lambda}_{cr}$, of MMPP_{cr} can be derived as follows

$$\mathbf{Q}_{cr} = \begin{bmatrix} -\varphi_{cr1} & \varphi_{cr1} \\ \varphi_{cr2} & -\varphi_{cr2} \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda}_{cr} = \begin{bmatrix} \lambda_{cr1} & 0 \\ 0 & \lambda_{cr2} \end{bmatrix} \tag{3.30}$$

where

$$\varphi_{cr1} = \frac{\tau}{2}\left(1 + \frac{1}{\sqrt{4e+1}}\right) \tag{3.31}$$

$$\varphi_{cr2} = \tau - \varphi_{cr1} \tag{3.32}$$

$$\lambda_{cr2} = \left(\frac{\lambda_{cr}\tau}{\varphi_{cr2}} - \frac{\psi}{\varphi_{cr1} - \varphi_{cr2}}\right)\frac{\varphi_{cr2}}{\varphi_{cr1} + \varphi_{cr2}} \tag{3.33}$$

$$\lambda_{cr1} = \frac{\psi}{\varphi_{cr1} - \varphi_{cr2}} + \lambda_{cr2} \tag{3.34}$$

The hot-spot traffic is non-uniformly distributed over network channels. In the presence of hot-spot traffic, the nodes located at j $(1 \le j \le d_{max})$ hops away from the hot-spot node can be depicted in Fig. 3.2. Therefore, the traffic characteristics on network channels with the same distance to the hot-spot node are identical. Given that there are C_j channels that are j hops away from the hot-spot node and $N_{\ge j} = \left(N - \sum_{i=0}^{j-1} N_i\right)$ nodes located more than j hops away from the hot-spot node. Therefore, the hot-spot traffic on network channels located at j hops

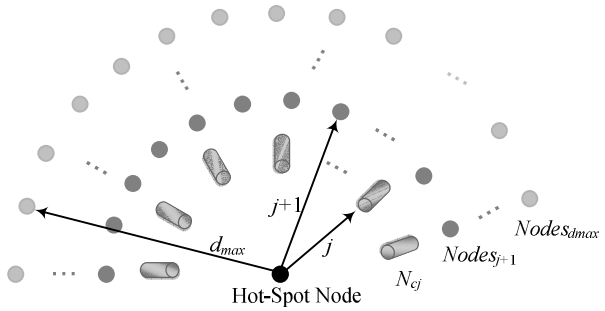

Figure 3.2: Number of channels located at j hops away from the hot-spot node and the number of nodes located at j hops farther away from the hot-spot node

away from the hot-spot node is f_{h_j} times of that generated by a source node. f_{h_j} can be expressed as

$$f_{h_j} = \frac{N_{\geq j}}{C_j}\delta = \frac{N - \sum_{i=0}^{j-1} N_i}{C_j}\delta \tag{3.35}$$

Adopting the similar method used in the derivation of the regular traffic on network channels, we can easily obtain the infinitesimal generator, $\mathbf{Q}_{ch_j}$, and rate matrix, $\mathbf{\Lambda}_{ch_j}$, of MMPP_{ch_j}, characterising the hot-spot traffic arriving at network channels located at j hops away from the hot-spot node. The superposition of these two types of traffic yields the loads at network channels that are j $(1 \leq j \leq d_{max})$ hops away from the hot-spot node, modelled by MMPP_{c_j} with the infinitesimal generator, $\mathbf{Q}_{c_j}$, and rate matrix, $\mathbf{\Lambda}_{c_j}$.

As adaptive routing distributes regular traffic evenly across the network channels and the torus topology is symmetrical, the service time experienced by a regular message on each channel is identical and is equal to its network latency, t_r. However, the presence of hot-spot traffic causes the service time to vary from one network channel to another depending on the locations with respect to the hot-spot

node. Considering both regular and hot-spot traffic with their corresponding weights yields the service time at network channels located at j $(1 \le j \le d_{max})$ hops away from the hot-spot node, t_{c_j}, as

$$t_{c_j} = \frac{\lambda_{cr}}{\lambda_{c_j}} t_r + \frac{\lambda_{ch_j}}{\lambda_{c_j}} t_{h_j} \tag{3.36}$$

where λ_{ch_j} and λ_{c_j} are the mean arrival rate of the hot-spot traffic and the superposed traffic on network channels located at j hops away from the hot-spot node and can be identified based on Eq. (3.19).

3.3.5 Message waiting time at network channels

To determine the waiting time, W_{c_j}, experienced by a message to acquire a channel located at j hops away from the hot-spot node, the physical channel can be treated as an MMPP/G/1 queueing system [31], where the arrival process is modelled by MMPP_{c_j} and the service time is t_{c_j}. Therefore, W_{c_j} can be given by

$$W_j = \frac{2\rho_j + \lambda_{c_j} t_{c_j}^{(2)} - 2t_{c_j}((1-\rho_j)\mathbf{g} + t_{c_j}\boldsymbol{\pi}_{c_j}\boldsymbol{\Lambda}_{c_j})(\mathbf{Q}_{c_j} + \mathbf{e}_{c_j}\boldsymbol{\pi}_{c_j})^{-1}\hat{\boldsymbol{\lambda}}}{2(1-\rho_j)} \tag{3.37}$$

$$W_{c_j} = \frac{1}{\rho_j}\left(W_j - \frac{1}{2}\lambda_{c_j} t_{c_j}^{(2)}\right) \tag{3.38}$$

where t_{c_j} and $t_{c_j}^{(2)}$ represent the first two moments of the service time on network channels located at j $(1 \le j \le d_{max})$ hops away from the hot-spot node; these two quantities can be obtained from the Laplace-Stieltjes transform, $F_{t_{c_j}}^{*}(s)$, of the service time on network channels [62]. Since the Laplace-Stieltjes transform of the

sum of independent random variables is equal to the product of their transforms. $F_{t_{c_j}}^{*}(s)$ can be expressed as follows

$$F_{t_{c_j}}^{*}(s) = \frac{\lambda_{cr}}{\lambda_{c_j}} F_{t_r}^{*}(s) \frac{\lambda_{ch_j}}{\lambda_{c_j}} F_{t_{h_j}}^{*}(s) \tag{3.39}$$

where $F_{t_r}^{*}(s)$ and $F_{t_{h_j}}^{*}(s)$ denote the Laplace-Stieltjes transforms of the network latencies for regular messages and j-hop hot-spot messages, respectively. Based on Eq. (3.39), $F_{t_r}^{*}(s)$ can be given by

$$F_{t_r}^{*}(s) = e^{-s(\omega+\bar{d}_r)} F_{b_r}^{*}(s) \tag{3.40}$$

where $F_{b_r}^{*}(s)$ represents the Laplace-Stieltjes transform of the blocking time experienced by the regular messages on network channels. Similarly, $F_{t_{h_j}}^{*}(s)$ can be written as

$$F_{t_{h_j}}^{*}(s) = e^{-s(\omega+j)} F_{b_{h_j}}^{*}(s) \tag{3.41}$$

where $F_{b_{h_j}}^{*}(s)$ denotes the Laplace-Stieltjes transform of the blocking time for j-hop hot-spot messages on network channels. The traffic intensity is $\rho_j = t_{c_j} \lambda_{c_j}$. $\hat{\boldsymbol{\lambda}} = \boldsymbol{\Lambda}_{c_j} \mathbf{e}_{c_j}$ and $\mathbf{e}_{c_j} = (1,1)^{\mathrm{T}}$. $\boldsymbol{\pi}_{c_j}$ is the steady-state vector of the MMPP_{c_j}. The algorithm for computing the matrix $\mathbf{g}$ can be found in [62].

Examining the expression of the blocking time experienced by a regular message on network channels, given by Eq. (3.12), we can find that it is infeasible to obtain the exact expression for its Laplace-Stieltjes transform, denoted by $F_{b_r}^{*}(s)$. Given that any distribution function can be approximated arbitrarily closely by a

series-parallel stage-type device [62]. For the analytical simplicity and versatility, the distribution of the message blocking time is approximated to an exponential distribution. Therefore, the probability density function of the blocking time for a regular message, $f_{b_r}(x)$, and its Laplace-Stieltjes transform, $F^*_{b_r}(s)$, can be written as

$$f_{b_r}(x) = \alpha e^{-\alpha x} \quad (\alpha > 0) \tag{3.42}$$

$$F^*_{b_r}(s) = \frac{\alpha}{\alpha + s} \tag{3.43}$$

where α is selected to match the mean blocking time experienced by a regular message on network channels. α can be given by

$$\alpha = \frac{1}{\sum_{i=1}^{\bar{d}_r} b_{r_i}} \tag{3.44}$$

Similarly, the Laplace-Stieltjes transform of the blocking time experienced by a *j*-hop hot-spot message on network channels, $F^*_{b_{h_j}}(s)$, can be given by

$$F^*_{b_{h_j}}(s) = \frac{\beta}{\beta + s} \tag{3.45}$$

where β is chosen to match the mean blocking time experienced by a *j*-hop hot-spot message on network channels.

3.3.6 Average degree of virtual channel multiplexing

The probability, $P_{v,j}$, that v $(0 \le v \le V)$ virtual channels at a given physical channel located at j $(1 \le j \le d_{max})$ hops away from the hot-spot node are busy can be determined using the probability that there are v packets in an MMPP/G/1 queueing

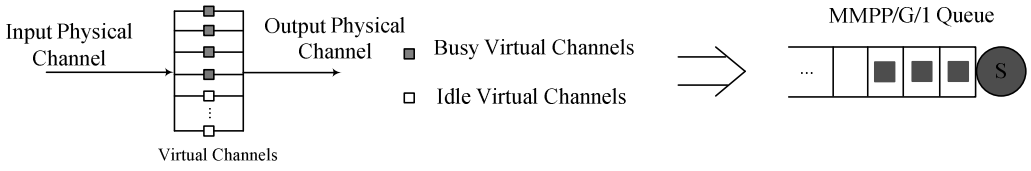

Figure 3.3: Calculation of the probability of busy virtual channels in each physical channel

system, as depicted in Fig. 3.3 [5]. Specifically, the probability that v virtual channels are busy, $0 \le v \le V-1$, corresponds to the probability that there are v packets in the queueing system, while the probability that all virtual channels are busy is the sum of the probabilities that there are v, $V \le v \le \infty$, packets in the queueing system. Therefore, $P_{v,j}$ can be given by [105]

$$P_{v,j} = \begin{cases} \boldsymbol{\pi}_{c_j}(\mathbf{I}_c - \mathbf{R}_j)\mathbf{R}_j^v \mathbf{e}_{c_j} & 0 \le v \le V-1 \\ \boldsymbol{\pi}_{c_j}\mathbf{R}_j^v \mathbf{e}_{c_j} & v = V \end{cases} \tag{3.46}$$

where the factor, $\mathbf{R}_j$, can be computed by solving the quadratic matrix equation

$$\mathbf{A}_j + \mathbf{R}_j\mathbf{B}_j + \mathbf{R}_j^2\mathbf{C}_j = 0 \tag{3.47}$$

with $\mathbf{A}_j = \boldsymbol{\Lambda}_{c_j}$, $\mathbf{B}_j = \mathbf{Q}_{c_j} - \boldsymbol{\Lambda}_{c_j} - \mathbf{I}_c / T_{c_j}$, and $\mathbf{C}_j = \mathbf{I}_c / T_{c_j}$. $\mathbf{I}_c$ is the identity matrix of size 2. A simple iterative method can be used to solve the above equation as follows [105]:

1) Initialising $\mathbf{R}_j = -\mathbf{A}_j / \mathbf{B}_j$;

2) At each iteration step, $\mathbf{R}_j = -(\mathbf{A}_j + \mathbf{R}_j^2\mathbf{C}_j)\mathbf{B}_j^{-1}$, until

$\left\| \mathbf{A}_j + \mathbf{R}_j\mathbf{B}_j + \mathbf{R}_j^2\mathbf{C}_j \right\| \le 10^{-10}$.

With the joint probability $P_{v,j}$, the average degree of virtual channel

multiplexing, $\overline{V}$, can be derived according to the method presented in [16].

3.3.7 Waiting time at the source node

Messages generated by each source node follow an MMPP_s and can enter the network through one of the V injection virtual channels with the equivalent probability $1/V$. Let MMPP_v denote the traffic arriving at an injection virtual channel at the source node. The infinitesimal generator, $\mathbf{Q}_v$, and rate matrix, $\mathbf{\Lambda}_v$, of MMPP_v can be determined from the splitting of MMPP_s with the splitting probability $1/V$. A regular message injected from a source node located at j $(1 \le j \le d_{max})$ hops away from the hot-spot node experiences a network latency of t_r given by Eq. (3.5), whereas a hot-spot message experiences the latency of t_{h_j} given by Eq. (3.6). Taking both the regular and hot-spot messages with their appropriate probabilities into consideration yields the service time for a message originating from a source node located at j hops away from the hot-spot node as

$$t_{s_j} = (1-\delta)t_r + \delta\, t_{h_j} \tag{3.48}$$

Based on Eq. (3.39), the Laplace-Stieltjes transform of the service time, t_{s_j} , can be written as

$$F^*_{t_{s_j}}(s) = (1-\delta)\delta F^*_{t_r}(s) F^*_{t_{h_j}}(s) \tag{3.49}$$

To determine the waiting time, W_{s_j} , experienced by a message at the source node that is located at j hops away from the hot-spot node, the local queue is modelled as an MMPP/G/1 queueing system. The derivation of W_{s_j} is similar to that

used for the determination of the waiting time, W_{c_j}, experienced by a message at network channels. Averaging over all possible values of j $(1 \le j \le d_{max})$ yields the mean waiting time at a given source node as

$$\overline{W}_s = \sum_{j=1}^{d_{max}} P_j W_{s_j} \tag{3.50}$$

3.4 Validation of the model

The accuracy of the analytical model is validated by means of a discrete event-driven simulator, operating at the flit level. Each simulation experiment was run until the network reached its steady state, which means that a further increase in the simulation time does not change the statistical results by any significant amount. The network cycle time in the simulator is defined as the transmission time of a single flit to traverse from one node to another. The message latency is defined as the mean amount of time from the generation of a message until the last data flit reaches PE at the destination node. Messages generated by each source node are modelled by an MMPP_s with the infinitesimal generator, $\mathbf{Q}_s$, and rate matrix, $\mathbf{\Lambda}_s$. The message destinations are non-uniformly distributed over the network nodes and are dispatched according to the hot-spot fraction, δ. The generated message has the probability, δ, to be directed to the hot-spot node, and the probability, $(1-\delta)$, of being evenly directed to all network nodes.

For each simulation experiment, the *batch means method* [17] is used to collect the statistics of performance metrics, i.e., the total number of messages are split into many batches and statistics are accumulated for these batches. Let b and $\bar{s}$ denote the number of batches and the size of each batch, respectively. Given a set of

observations $\{X_0, X_1, \cdots, X_{\bar{s}b-1}\}$ from a single simulation run, where X_i $(0 \le i \le \bar{s}b-1)$ is the message latency of the *i*-th message. The message latency, L_k, accumulated for the *k*-th batch can be determined by

$$L_k = \frac{1}{\bar{s}} \sum_{\ell=0}^{\bar{s}-1} X_{\bar{s}k+\ell} \tag{3.51}$$

where $0 \le k \le b-1$. Then the mean message latency can be obtained as the mean of the batch means as follows

$$\overline{L} = \frac{1}{b} \sum_{k=0}^{b-1} L_k \tag{3.52}$$

In our simulation experiments, to make sure the simulation has reached the steady state, 8,000 messages were discarded for the warm-up phase. The message latency is gathered for a total number of 80,000 messages, where these messages are split into 10 batches and the size of each batch is 8,000 messages. Following Eqs. (3.51) and (3.52), the mean message latency can be determined. Since each sample in the batch means method is an average over many of the original samples, the variance between batch means is greatly reduced, which in turn decreases the standard deviation of the measurement performance metrics. This leads to the better confidence in the performance results.

Extensive simulation experiments have been performed to validate the model for various combinations of network sizes, message lengths, number of virtual channels per physical channel, different MMPP_s traffic inputs, and various hot-spot fractions. However, for the sake of specific illustration and without loss of generality, latency results are presented for the following cases:

- Network size N = 12×12 and 16×16 nodes;

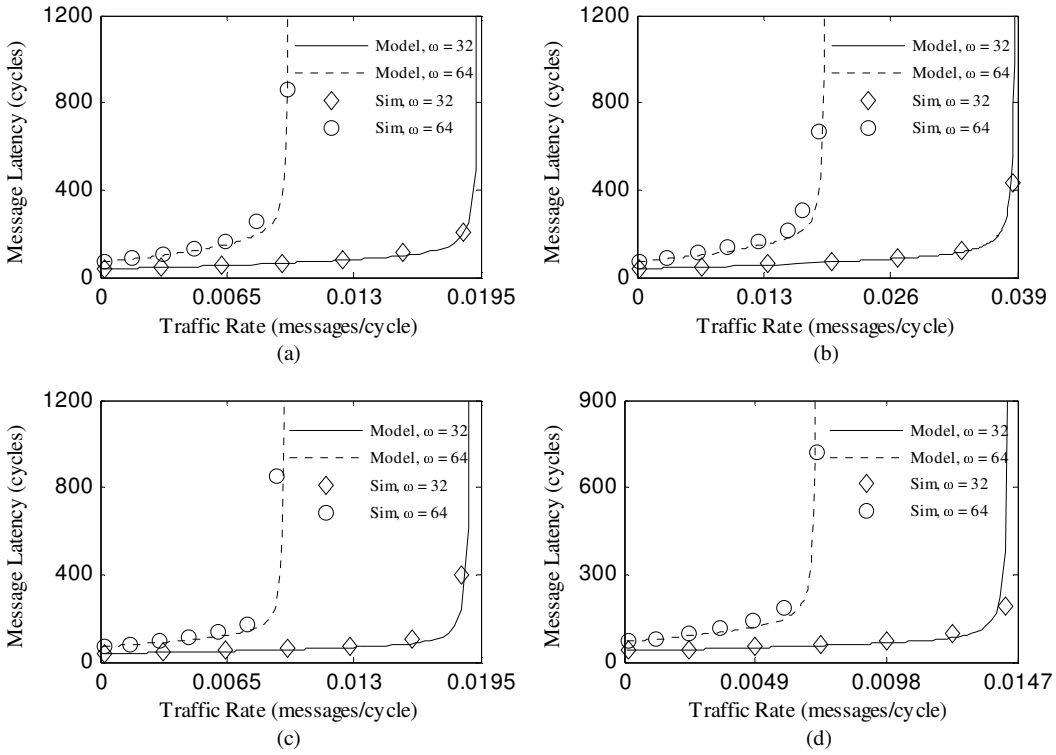

Figure 3.4: Latency predicted by the model against simulation experience in the 12×12 torus with: (a) $V = 4$, $\varphi_{s1} = 0.5$, $\varphi_{s2} = 0.5$, $\delta = 0.05$, (b) $V = 5$, $\varphi_{s1} = 0.6$, $\varphi_{s2} = 0.2$, $\delta = 0.05$, (c) $V = 6$, $\varphi_{s1} = 0.8$, $\varphi_{s2} = 0.4$, $\delta = 0.1$ and (d) $V = 7$, $\varphi_{s1} = 0.9$, $\varphi_{s2} = 0.7$, $\delta = 0.1$

- Message length $\omega = 32$ and 64 flits;
- Number of virtual channels is $V = 4$, 5, 6 and 7 per physical channel;
- Parameters φ_{s1} and φ_{s2} in the infinitesimal generator $\mathbf{Q}_s$ of MMPP$_s$ are set as follows:

$$\begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}, \begin{bmatrix} -0.6 & 0.6 \\ 0.2 & -0.2 \end{bmatrix}, \begin{bmatrix} -0.8 & 0.8 \\ 0.4 & -0.4 \end{bmatrix}, \begin{bmatrix} -0.9 & 0.9 \\ 0.7 & -0.7 \end{bmatrix},$$

$$\begin{bmatrix} -0.8 & 0.8 \\ 0.6 & -0.6 \end{bmatrix}, \begin{bmatrix} -0.6 & 0.6 \\ 0.4 & -0.4 \end{bmatrix}, \begin{bmatrix} -0.7 & 0.7 \\ 0.35 & -0.35 \end{bmatrix} \text{ and } \begin{bmatrix} -0.9 & 0.9 \\ 0.3 & -0.3 \end{bmatrix};$$

- Hot spot fraction $\delta = 0.05$ and 0.1.

The infinitesimal generator, $\mathbf{Q}_s$, of MMPP$_s$ and the hot spot fraction, δ,

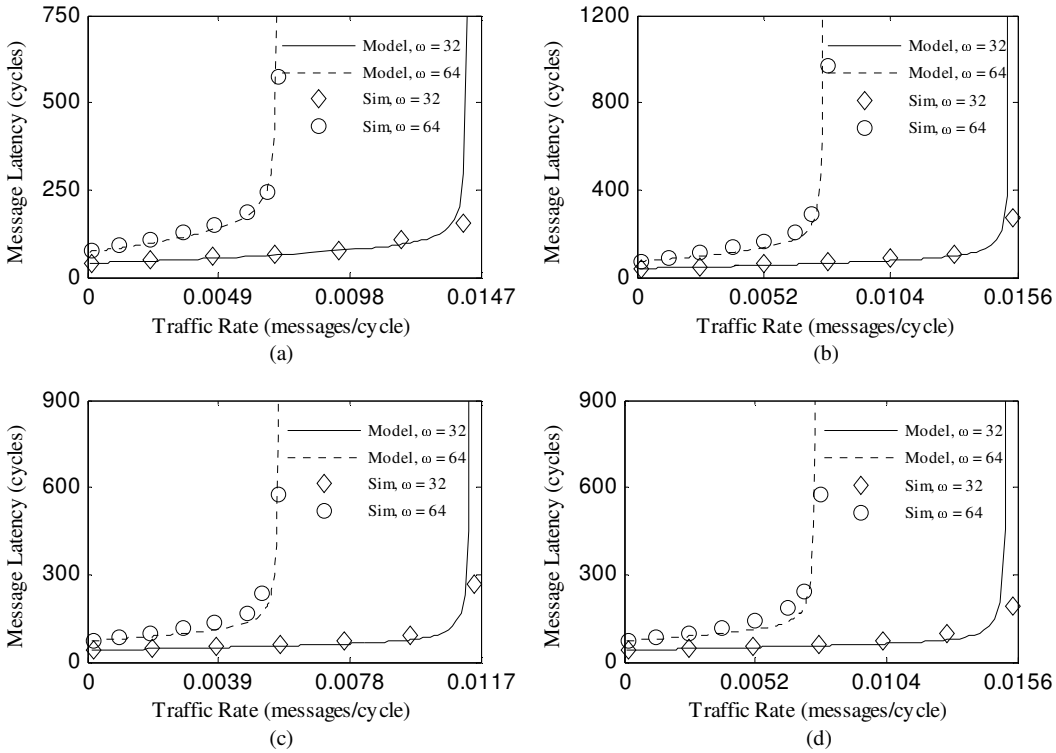

Figure 3.5: Latency predicted by the model against simulation experience in the 16×16 torus with: (a) $V = 4$, $\varphi_{s1} = 0.8$, $\varphi_{s2} = 0.6$, $\delta = 0.05$, (b) $V = 5$, $\varphi_{s1} = 0.6$, $\varphi_{s2} = 0.4$, $\delta = 0.05$, (c) $V = 6$, $\varphi_{s1} = 0.7$, $\varphi_{s2} = 0.35$, $\delta = 0.1$ and (d) $V = 7$, $\varphi_{s1} = 0.9$, $\varphi_{s2} = 0.3$, $\delta = 0.1$

represent different degrees of traffic burstiness in the temporal domain and spatial domain, respectively. Figs. 3.4 and 3.5 depict the performance results for the message latency predicted by the analytical model plotted against those obtained from the simulator as a function of the traffic rate in the 12×12 and 16×16 torus. In these figures, the horizontal axis represents the traffic rate, λ_{s1}, at which a node injects messages into the network when the MMPP_s is at state 1, while the vertical axis denotes the message latency. For the sake of clarity of the figures, the arrival rate, λ_{s2}, at state 2 is deliberately set to zero; otherwise three-dimensional graphs will be used to represent the results. These figures reveal that the message latency obtained from the analytical model closely match those obtained from the simulation.

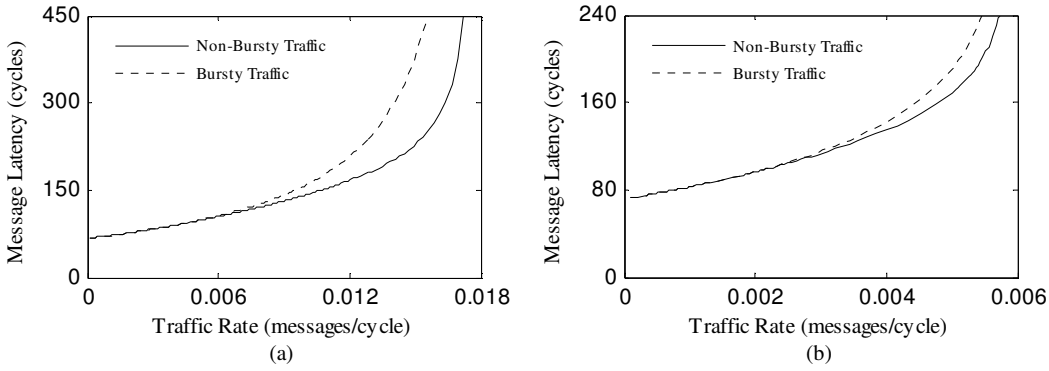

Figure 3.6: Message latency predicted by the above model and the model under the Poisson assumption with $\omega = 64$, $V = 4$, $\varphi_{s1} = 0.006$, $\varphi_{s2} = 0.006$, $\delta = 0.05$ in (a) 8×8 and (b) 16×16 torus network

The tractability and accuracy of the model make it a practical and cost-effective tool to gain insight into the performance of INs in multi-computer systems in the presence of bursty traffic with hot-spot destinations.

3.5 Performance analysis

Having validated the accuracy of the analytical model, we use it as a cost-effective tool to investigate the effects of bursty traffic with hot-spot destinations on the performance of INs in multi-computer systems.

3.5.1 The impact of bursty traffic on network performance

Fig. 3.6 shows the impact of bursty traffic on network performance through comparing the message latency as a function of the mean arrival rate when the network is subject to the bursty MMPP_s and non-bursty Poisson process, respectively. The mean arrival rate, $\overline{\lambda}_s$, of MMPP_s can be obtained by virtue of Eq. (2.11). As depicted in Fig. 3.6(a), the model based on the non-bursty Poisson process seriously underestimates the message latency when the network is subject to the

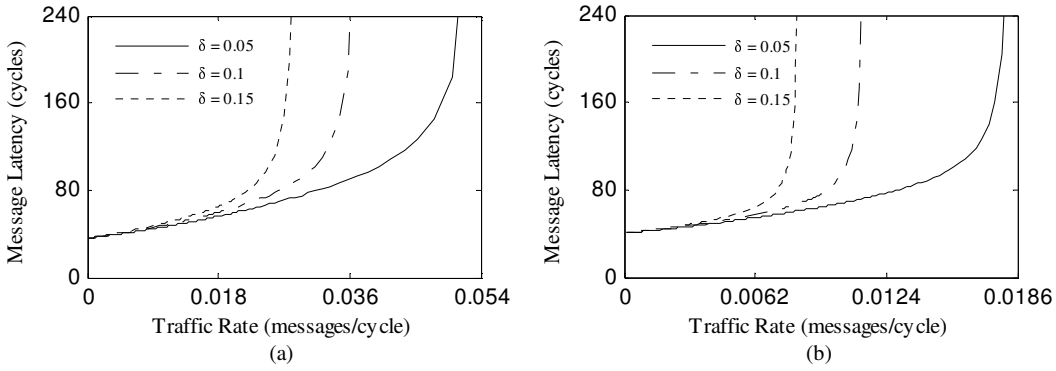

Figure 3.7: Message latency predicted by the above model with $\omega = 32$, $V = 4$, $\varphi_{s1} = 0.9$, $\varphi_{s2} = 0.45$ in (a) 8×8 and (b) 16×16 torus network

bursty MMPP_s. In contrast, the derived analytical model manages to predict the increase in message latency and decrease in the maximum network throughput in the presence of bursty traffic. Moreover, such a phenomenon is more apparent when the traffic is under moderate and heavy loads. The same result can be found in the network with larger sizes as shown in Fig. 3.6(b).

3.5.2 The impact of hot-spot destinations on network performance

Fig. 3.7 shows the message latency in a torus with the network size of 8×8 and 16×16, respectively, the degree of traffic burstiness imposed by the parameters of MMPP_s (i.e., $\varphi_{s1} = 0.9$, $\varphi_{s2} = 0.45$) and varying hot-spot fractions (i.e., $\delta = 0.05$, 0.1, and 0.2). The analytical results obtained from the model demonstrate that the network with a larger size suffers the higher message latency. That is because the message distance increases as the network size rises. Moreover, the increase of the hot-spot fraction degrades the network performance since the network saturates at a lower traffic rate. That is because the hot-spot traffic can cause the higher traffic

loads on network channels located closer to the hot-spot node. As the hot-spot fraction increases, these channels become overloaded quickly. Moreover, under the same hot-spot fraction, the large-size network becomes saturated more rapidly since the number of nodes which send their generated hot-spot messages to the hot-spot destination node increases.

These results emphasise the importance of using the more realistic traffic models than those developed under the non-bursty Poisson process with uniformly distributed message destinations to evaluate and optimise the network performance. Moreover, the bursty traffic with hot-spot destinations can deteriorate the network performance seriously, stressing the great need for the new network architectures that can efficiently support such a traffic pattern.

3.6 Conclusions

The traffic patterns in INs reveal the bursty nature in the both temporal domain and spatial domain. To capture the characteristics of realistic traffic patterns, this chapter has developed a new analytical performance model to investigate the performance of INs in multi-computer systems in the presence of bursty traffic with hot-spot destinations. The comparison between the analytical results with those obtained from extensive simulation experiments has shown that the analytical model possesses a good degree of accuracy for predicting the network performance under different design alternatives and various traffic conditions. The analytical results have shown that the network performance degrades considerably under this type of traffic pattern. Moreover, the results have stressed the importance of using the realistic traffic models to investigate and optimise the performance of INs in multi-computer systems.

Chapter 4

Performance Modelling of Interconnection Networks in Multi-cluster Systems under Bursty Traffic with Communication Locality

Analytical models for INs in cluster systems have been widely reported based on the simplified assumptions that the message arrivals follow the non-bursty Poisson process and message destinations are uniformly distributed over all network nodes [5, 14, 21, 56, 73, 91]. However, the assumption of uniform destination distribution is not always realistic in practice and the communication locality, a typical example of non-uniform destination distributions, has been shown to be an important phenomenon exhibited by many applications in cluster-based systems [2, 61, 115]. A number of measurement studies [19, 111] have revealed that the traffic generated by many real-world applications exhibits a high degree of burstiness (i.e. time-varying arrival rates) and possesses correlations in the number of message arrivals. Therefore, the traditional models with the simplified assumptions cannot capture the characteristics of realistic network traffic. It is very important and timely to re-evaluate the performance of networks in cluster systems under more realistic traffic models before practical implementations show their potential faults. To this end, this

chapter presents a new analytical model to investigate the performance of INs in multi-cluster systems in the presence of bursty traffic with communication locality. The validity of the model is demonstrated by comparing analytical results to those obtained through extensive simulation experiments of an actual system. The model is then used to evaluate the effects of bursty traffic with communication locality on the performance of INs in multi-cluster systems.

The rest of this chapter is organised as follows. Section 4.1 shows preliminaries including the architecture of multi-cluster systems, the topology of *m*-port *n*-tree, the switching method and routing strategy for message flow control, and the communication locality. The assumptions and notations adopted in this chapter are listed in Section 4.2. Section 4.3 derives the analytical model to investigate the performance of INs in multi-cluster systems. The results obtained from the analytical model are validated by extensive simulation experiments in Section 4.4. Section 4.5 carries out performance analysis. Finally, Section 4.6 concludes this study.

4.1 Preliminaries

4.1.1 The architecture of the multi-cluster systems

This chapter studies a multi-cluster system consisting of C clusters, each of which contains N nodes (PNs) with their own local memories (LMs) in each cluster, as depicted in Fig. 4.1. Each cluster has two communication networks: an intrA-Communication Network (ACN) used for the purpose of message passing between nodes within the same cluster and an intEr-Communication Network (ECN) used to transmit messages between clusters and manage the entire system [56]. The clusters are connected to each other by an Integrated Communication Network (ICN). The

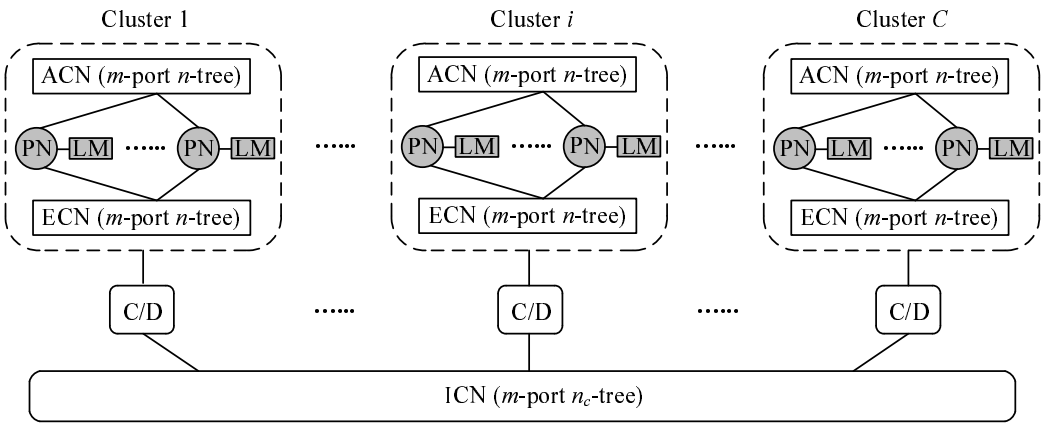

Figure 4.1: The architecture of multi-cluster systems

ECN and ICN are connected by a set of Concentrators/Dispatchers (C/Ds) [17], which are used to combine message traffic from/to one cluster to/from other clusters. Each communication network (i.e., ACN, ECN, and ICN) is constructed in the *m*-port *n*-tree topology. The specification of *m*-port *n*-tree will be presented in Section 4.1.2. It is worth noting that the ECN can be accessed directly by the node in each cluster without going through the ACN.

The construction of ACN and ECN based on *m*-port *n*-tree is straightforward. In the ICN, each cluster can be viewed again as the node of the tree. Let $(n_c + 1)$ denote the height of the tree in the ICN. Since there are C clusters in the system, n_c can be obtained by solving the equation $C = 2(m/2)^{n_c}$ as

$$n_c = \left\lceil \frac{\log_2 C - 1}{\log_2 m - 1} \right\rceil \tag{4.1}$$

4.1.2 Topology of *m*-port *n*-tree

Fat-tree is an efficient network topology to provide high-performance and low-latency infrastructure and has increased in popularity to construct the INs in cluster systems [37, 116]. The fat-tree topology and its variants are widely adopted in

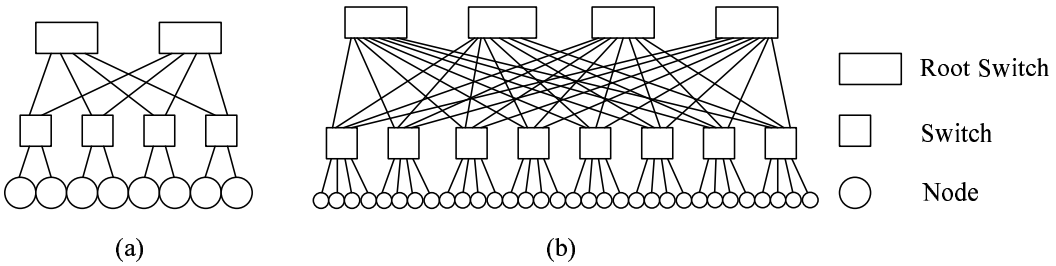

Figure 4.2: Some instances of *m*-port *n*-tree topology: (a) 4-port 2-tree and (b) 8-port 2-tree

practice. For instance, owing to its simple physical implementation and high-performance, the *m*-port *n*-tree, a typical example of fat-tree topologies, has been used to construct InfiniBand architecture (IBA) [127]. In addition, the *m*-port *n*-tree topology has full bisectional bandwidth and can offer a rich connectivity among nodes, making it possible to maintain paths between all source and destination nodes [36]. Fig. 4.2 depicts some instances of *m*-port *n*-tree topology. The *m*-port *n*-tree consists of N nodes and S switches (including the root switches) [69]. N and S can be given by

$$N = 2\left(\frac{m}{2}\right)^n \tag{4.2}$$

$$N_{sw} = (2n-1)\left(\frac{m}{2}\right)^{n-1} \tag{4.3}$$

In this topology, each network switch has m communication ports that are connected with other switches or nodes. The height of the tree is $(n+1)$. The switches (except for root switches) use half of the ports to connect with their descendants or nodes, and use the other half to connect with their ancestors. The root switches use all communication ports for the connection with their descendants or nodes.

A node in an *m*-port *n*-tree can be labelled as $PN(p_0, p_1, \cdots, p_{n-1})$, where $i \in \{0,1,\cdots,m-1\} \times \{0,1,\cdots,(m/2)-1\}^{n-1}$. The switch in *m*-port *n*-tree can be labelled as $SW < s = s_0 s_1 \cdots s_{n-2}, \ell >$, where $\ell \in \{0,1,\cdots,n-1\}$ is the level of the switch and

$$s \in \begin{cases} \{0,1,\cdots,(m/2)-1\}^{n-1} & \ell = 0 \\ \{0,1,\cdots,m-1\} \times \{0,1,\cdots,(m/2)-1\}^{n-2} & \ell \in \{1,2,\cdots,n-1\} \end{cases}$$

Let $SW < s, \ell >_k$ represent the *k*-th port of $SW < s, \ell >$, where $k = 0,1,\cdots,m-1$. For switches $SW < s, \ell >$ and $SW < s', \ell' >$, ports $SW < s, \ell >_k$ and $SW < s', \ell' >_{k'}$ are connected by a channel if and only if $\ell' = \ell + 1$, $s_0 s_1 \cdots s_{n-3} = s'_0 s'_1 \cdots s'_{\ell-1} s'_{\ell+1} \cdots s'_{n-2}$, $k = s'_\ell$, and $k' = s_{n-2} + (m/2)$. For switch $SW < s, n-1 >$, port $SW < s, n-1 >_k$ is connected to node $PN(p_0, p_1, \cdots, p_{n-1})$ if and only if $s_0 s_1 \cdots s_{n-2} = p_0 p_1 \cdots p_{n-2}$ and $k = p_{n-1}$.

4.1.3 Switching method and routing strategy

Owing to the attractive properties of wormhole switching in INs as specified in Section 3.1.2, we employ this switching method for the message flow control in this chapter. Due to the operation of wormhole switching, deadlock-free routing is required to avoid the deadlock occurrence in the network.

In deterministic routing [23], a message traverses a fixed path between source and destination, which simplifies the implementation, avoids message deadlock, and guarantees an in-order delivery. Very recently, Gomez *et al.* [36] have shown that deterministic routing can achieve a similar, and in some scenarios higher, level of performance than adaptive routing in fat-trees. Therefore, a deterministic routing based on the Up*/Down* algorithm [104] is adopted. In this routing strategy, a

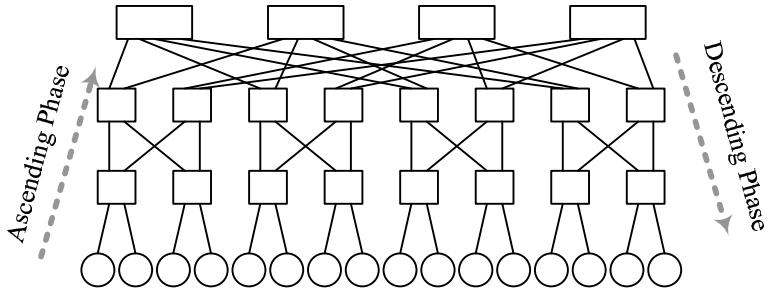

Figure 4.3: The deterministic routing in *m*-port *n*-tree topology

message traverses an *ascending phase* to get to a nearest common ancestor (NCA), followed by a *descending phase*, as depicted in Fig. 4.3. It has been shown [56] that this routing algorithm is able to perform a balanced traffic distribution over the whole network.

4.1.4 Non-uniform message destination distributions

The communication locality, a typical example of non-uniform destination distributions, has been shown to be an important phenomenon exhibited by many applications in cluster-based systems [2, 61, 115]. The communication locality exists when the likelihood of communications to various nodes decrease with the distance, which implies that the mean inter-node distance with communication locality is smaller than that with the uniform destination distribution [23]. As a result, each message consumes fewer resources and thus reduces contention. Agarwal [2] showed that the communication locality exhibited by the application can significantly improve both network throughput and message latency in communication networks. Larkins *et al.* [65] recently employed a global address space view of distributed tree data structure to improve the locality in high-performance parallel applications and thus yield good system performance.

4.2 Assumptions and notations

The model is based on the following assumptions. Assumptions (a) and (b) are used to generate the bursty traffic with communication locality, and the other assumptions are commonly accepted in the related studies [5, 14, 21, 54, 56, 73, 81, 82, 91, 120]. A brief summary of key notations used in the derivation of the model is listed in Table 4.1.

a) The arrivals of messages generated by each source node are bursty and follow a two-state MMPP$_s$ (the subscript s denoting the traffic generated by the source node). The MMPP$_s$ can be characterised by the infinitesimal generator, $\mathbf{Q}_s$, of the underlying Markov chain and rate matrix, $\mathbf{\Lambda}_s$. $\mathbf{Q}_s$ and $\mathbf{\Lambda}_s$ can be given by

$$\mathbf{Q}_s = \begin{bmatrix} -\varphi_{s1} & \varphi_{s1} \\ \varphi_{s2} & -\varphi_{s2} \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda}_s = \begin{bmatrix} \lambda_{s1} & 0 \\ 0 & \lambda_{s2} \end{bmatrix} \tag{4.4}$$

b) Messages generated from the source node can be sent to the ECN with the probability, ξ, and to the ACN with the probability, $(1-\xi)$. Let us refer to the messages sent to the ECN and the ACN as inter-cluster messages and intra-cluster messages, respectively. Since inter-cluster messages are directed to a node in any other clusters with the equivalent probability, this probability is defined as the degree of inter-cluster locality, η_{out}, i.e., $\eta_{out} = \xi$ [54]. Messages are directed to a node within the cluster with intra-cluster locality, η_{in};

c) Message length is equal to ω flits. The length of each flit is L_f bytes.

Table 4.1: Key notations used in the derivation of the model in Chapter 4

$\overline{Latency}$	mean message latency for the multi-cluster system

Lat_{in}	message latency for intra-cluster communication networks
Lat_{out}	message latency for inter-cluster communication networks
η_{in} , η_{out}	intra-cluster and inter-cluster locality
ξ	the probability that a message generated from the source node is sent to the ECN
P_j	the probability of an inter-cluster message traversing $2j$ channels to reach its destination
$P_{j,\eta_{in}}$	the probability of an intra-cluster message traversing $2j$ channels to reach its destination
$\overline{d}_{in}$, $\overline{d}_{out}$	the mean message distance for intra-cluster messages and inter-cluster message
$\mathbf{Q}_s$, $\mathbf{\Lambda}_s$	parameter matrices of MMPP_s to model the traffic generated by the source node
$\mathbf{Q}_{cA}$, $\mathbf{\Lambda}_{cA}$	parameter matrices of MMPP_{cA} to model the traffic arriving at a given network channel in the ACN
$\mathbf{Q}_{sA}$, $\mathbf{\Lambda}_{sA}$	parameter matrices of MMPP_{sA} to model the traffic arriving at an injection channel at the source node in the ACN
MMPP_{cE}	the traffic arriving at a given network channel in the ECN
MMPP_{cI}	the traffic arriving at a given network channel in the ICN
T_{in} , T_{out}	mean network latency for intra-cluster communication networks and inter-cluster communication networks
W_{in} , W_{out}	waiting time at the source in intra-cluster networks and inter-cluster networks
W_c , W_d	the waiting time at the Concentrator and Dispatcher
τ_{in}, τ_{out}	the mean time for the tail to reach the destination in intra-cluster communication networks and inter-cluster communication networks
T_j, T_{p+q}	the network latency of a $2j$-channel intra-cluster message and a $2(p+q)$-channel inter-cluster message
$T_{k,j}$, $T_{k,p+q}$	the service time experienced by the $2j$-channel intra-cluster message and $2(p+q)$-channel inter-cluster message on a channel at the stage k

t_{cn}^{A}, t_{cn}^{E}	the time for a single flit to transmit on a node-to-switch (or switch-to-node) connection in the ACN and ECN
t_{cs}^{A}, t_{cs}^{E}, t_{cs}^{I}	the time for a flit to transmit on a switch-to-switch connection in the ACN, ECN and ICN
$Wb_{k,j}$	the blocking time that a 2*j*-channel message acquires a channel at stage k in intra-cluster networks
$Wb_{k,p+q}$	the blocking time that a $2(p+q)$ -channel message acquires a channel at stage k in inter-cluster networks
$Pb_{k,j}$	the probability that a 2*j*-channel intra-cluster message is blocked at stage k
$Wc_{k,j}$	the waiting time experienced by a 2*j*-channel intra-cluster message to acquire a channel when blocking occurs at stage k
$F_{k,j}^{*}(s)$	Laplace-Stieltjes transform of message service time on network channels in ACN
$W_{k,j}^{*}(s)$	Laplace-Stieltjes transform of the blocking time experienced by a 2*j*-channel message on network channels at stage k

4.3 The analytical model

The mean message latency for the multi-cluster system can be obtained by the weighted sum of the message latency in intra-cluster communication networks (i.e., ACN), Lat_{in}, and inter-cluster communication networks (i.e., ECN and ICN), Lat_{out}, as follows

$$\overline{Latency} = (1-\xi)Lat_{in} + \xi Lat_{out} \tag{4.5}$$

In what follows, the traffic characteristics on network channels in the ACN, ECN, and ICN will first be determined, and then the message latency in intra-cluster and inter-cluster communication networks will be derived.

4.3.1 Traffic characteristics on channels in intra-cluster and

inter-cluster communication networks

The traffic pattern, in the presence of communication locality, mainly affects the average number of channels that a message need cross to reach its destination, commonly known as mean message distance. In an *m*-port *n*-tree, the probability of a message traversing $2j$ channels (j channels in ascending phase and j channels in descending phase) to reach its destination is P_j. Different choices of P_j could lead to different distributions of message destinations and, consequently, different mean message distances. As described in Assumption (b), inter-cluster messages are destined to a node in any other clusters with the equivalent probability. As shown in Fig. 4.2, the number of nodes at distance $2n$ is $(m/2)^{n-1}(m-1)$ in the *m*-port *n*-tree [69], therefore, P_j can be given by

$$P_j = \begin{cases} \dfrac{(m-1)\left(\dfrac{m}{2}\right)^{j-1}}{N-1} & j = n \\ \dfrac{\left(\dfrac{m}{2}-1\right)\left(\dfrac{m}{2}\right)^{j-1}}{N-1} & 1 \le j < n \end{cases} \tag{4.6}$$

Intra-cluster messages are destined to a node in each cluster with the same locality, η_{in}. The model proposed by Agrawal [2] is adopted for the intra-cluster locality in this chapter. For a given node, message destinations are chosen randomly from $\eta_{in}N$ nodes centred at the source node. In particular, destinations chosen uniformly over all the nodes in the network correspond to $\eta_{in} = 1$. With this locality model, the probability of an intra-cluster message traversing $2j$ channels to reach its destination with the communication locality, η_{in}, can be derived based on Eq. (4.6) as follows:

$$P_{j,\eta_{in}} = \begin{cases} \dfrac{(m-1)\left(\dfrac{m}{2}\right)^{j-1}}{\alpha(\eta_{in},m,n)}\eta_{in}^{j} & j = n \\ \dfrac{\left(\dfrac{m}{2}-1\right)\left(\dfrac{m}{2}\right)^{j-1}}{\alpha(\eta_{in},m,n)}\eta_{in}^{j} & 1 \le j < n \end{cases} \tag{4.7}$$

where $\alpha(\eta_{in},m,n)$ is a normalising factor and is chosen such that $\sum_{j=1}^{n} P_{j,\eta_{in}} = 1$. Therefore,

$$\alpha(\eta_{in},m,n) = \frac{\left(\dfrac{\eta_{in}m}{2}\right)^{n}\left(m-1-\dfrac{1}{\eta_{in}}\right)-\dfrac{m}{2}+1}{\dfrac{m}{2}-\dfrac{1}{\eta_{in}}}$$

The mean message distance for intra-cluster messages, $\bar{d}_{in}$, and inter-cluster messages, $\bar{d}_{out}$, can be obtained by averaging over all the possible values of j as follows:

$$\bar{d}_{in} = \sum_{j=1}^{n} 2jP_{j,\eta_{in}} \tag{4.8}$$

$$\bar{d}_{out} = \sum_{j=1}^{n} 2jP_{j} \tag{4.9}$$

Recall that messages generated from each source node has the probability, $(1-\xi)$, to be sent to the ACN with the mean message distance $\bar{d}_{in}$, given by Eq. (4.8). Since there are N nodes in each cluster and the messages generated from these nodes are sent to $4nN$ channels in the network, the traffic arriving at a network channel in the ACN is f_A times of the traffic generated by a source node. f_A can be expressed as

$$f_A = \frac{N(1-\xi)\bar{d}_{in}}{4nN} = \frac{(1-\xi)}{4n}\bar{d}_{in} \tag{4.10}$$

In addition, messages generated by the source node have the probability, ξ, to be sent to the ECN in the source cluster and through the ICN to reach its destination in the ECN in the destination cluster with the mean message distance $\bar{d}_{out(E)}$ and $\bar{d}_{out(I)}$ in the ECN and ICN, respectively. $\bar{d}_{out(E)}$ and $\bar{d}_{out(I)}$ can be obtained by Eq. (4.9). Similarly, the traffic arriving at a network channel in the ECN and the ICN is f_E and f_I times, respectively, of the traffic generated by a source node. Based on Eq. (4.10), f_E and f_I can be expressed as

$$f_E = \frac{(N\xi + N\xi)\bar{d}_{out(E)}}{4nN} = \frac{2\xi}{4n}\bar{d}_{out(E)} \tag{4.11}$$

$$f_I = \frac{CN\xi\bar{d}_{out(I)}}{4n_cC} = \frac{N\xi}{4n_c}\bar{d}_{out(I)} \tag{4.12}$$

where n_c is given by Eq. (4.1).

Since the splitting and superposition of multiple MMPPs give rise to a new MMPP [31], let MMPP_{cA}, MMPP_{cE}, and MMPP_{cI} denote the traffic arriving at a given network channel in the ACN, ECN, and ICN, respectively. Based on Eqs. (3.17)-(3.34), the parameter matrices of MMPP_{cA}, MMPP_{cE}, and MMPP_{cI} can be easily obtained.

4.3.2 Message latency in intra-cluster communication networks

In this section, we will derive the message latency in intra-cluster communication networks (i.e., ACN), Lat_{in}, which consists of three parts: a) network latency, T_{in},

that is the time for a message to cross the network; b) the waiting time at the source node, W_{in}; and c) the time for the tail to reach the destination, τ_{in}. Therefore, Lat_{in} can be expressed as

$$Lat_{in} = T_{in} + W_{in} + \tau_{in} \tag{4.13}$$

The quantities, T_{in}, W_{in}, and τ_{in}, will be derived in the subsequent sections.

4.3.2.1 Network latency for intra-cluster messages

Since an intra-cluster message may cross a different number of channels to reach its destination, we take into account the network latency of a 2*j*-channel message (i.e., a message need traverse $2j$ channels to reach its destination) as T_j. For the sake of illustration, the network stages are numbered as follows: the network stage numbering is based on the location of switches between the source and destination. The numbering starts from the stage next to the source node (stage 0) and goes up as we get closer to the destination node (stage $K-1$). In an *m*-port *n*-tree, the number of stages to be crossed for a 2*j*-channel message is $K = 2j-1$. Since messages are transferred to the local nodes as soon as they arrive at their destinations, we first consider the service time experienced by a message at the final stage and then continue the analysis backward to the first stage. The service time of a channel experienced by a 2*j*-channel message at stage $K-1$, $T_{K-1,j}$, can be found as

$$T_{K-1,j} = \omega t_{cn}^{A} \tag{4.14}$$

where ω is the message length in flits and t_{cn}^{A} is the time for a single flit to transmit on a node-to-switch (or switch-to-node) connection in the ACN and can be computed as

$$t_{cn}^{A} = 0.5\theta_{n}^{A} + L_f \gamma_{n}^{A} \tag{4.15}$$

where θ_{n}^{A} and γ_{n}^{A} are the network latency and the transmission time of one byte (inverse of bandwidth) in the ACN.

The service time at the internal stages k $(0 \le k \le K-2)$ might be more, since a channel would be idle when the channels of subsequent stages are busy. The service time experienced by messages on network channels at the internal stages can be obtained by the sum of the actual message transmission time and the delay due to blocking at the subsequent stages. Thus, the service time experienced by the 2*j*-channel message, $T_{k,j}$, on a channel at the internal stages can be found as

$$T_{k,j} = \omega t_{cs}^{A} + \sum_{\ell=k+1}^{K-1} Wb_{\ell,j}, \quad 0 \le k \le K-2 \tag{4.16}$$

where t_{cs}^{A} is the time for a flit to transmit on a switch-to-switch connection in the ACN and $Wb_{k,j}$ is the blocking time that a message acquires a channel at stage k. t_{cs}^{A} can be calculated as

$$t_{cs}^{A} = \theta_{s}^{A} + L_f \gamma_{n}^{A} \tag{4.17}$$

where θ_{s}^{A} denotes the switch latency in the ACN. $Wb_{k,j}$ can be determined by the probability that a message is blocked at this stage, $Pb_{k,j}$, and the waiting time experienced by the message to acquire a channel when blocking occurs, $Wc_{k,j}$. Therefore, $Wb_{k,j}$ can be expressed as

$$Wb_{k,j} = Pb_{k,j} Wc_{k,j} \tag{4.18}$$

To compute $Pb_{k,j}$, let us first calculate the joint probability, $P_{a,b}$ $(0 \le a \le 1)$

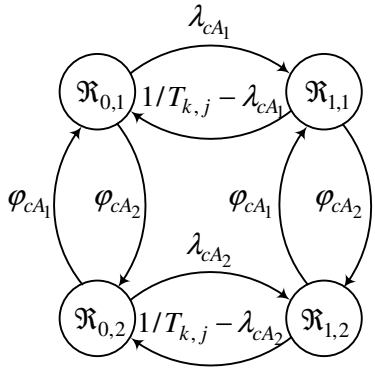

Figure 4.4: Markov chain to compute the blocking probability at network channels

and $(b = 1,2)$, that the channel is idle or busy and the underlying Markov chain of MMPP_{cA}, characterising the traffic on network channels in the ACN, is at the state b. $P_{a,b}$ can be determined using a bi-variate Markov chain as shown in Fig. 4.4, where state $\Re_{a,b}$ corresponds to the case where the channel is idle $(a = 0)$ or the channel is busy $(a = 1)$, and the MMPP_{cA} is at state b. The transition rate out of state $\Re_{a,b}$ to $\Re_{a+1,b}$ is λ_{cA_b}, where λ_{cA_b} is the traffic rate on network channels in the ACN when MMPP_{cA} is at state b; while the rate from $\Re_{a+1,b}$ to $\Re_{a,b}$ is $1/T_{k,j} - \lambda_{cA_b}$. The reduction of λ_{cA_b} is used to account for the arrival of messages while a channel is in this state [16]. The transition rates, φ_{cA_b}, are given by the infinitesimal generator $\mathbf{Q}_{cA}$, while the arrival rates λ_{cA_b} can be found from rate matrix $\mathbf{\Lambda}_{cA}$. According to Fig. 4.4, the model yields the following steady state equations

$$\left.\begin{aligned}
&(\lambda_{cA_1} + \varphi_{cA_2})P_{0,1} = \varphi_{cA_1}P_{0,2} + (1/T_{k,j} - \lambda_{cA_1})P_{1,1} \\
&(\lambda_{cA_2} + \varphi_{cA_1})P_{0,2} = \varphi_{cA_2}P_{0,1} + (1/T_{k,j} - \lambda_{cA_2})P_{1,2} \\
&(1/T_{k,j} - \lambda_{cA_1} + \varphi_{cA_2})P_{1,1} = \lambda_{cA_1}P_{0,1} + \varphi_{cA_1}P_{1,2} \\
&(1/T_{k,j} - \lambda_{cA_2} + \varphi_{cA_1})P_{1,2} = \lambda_{cA_2}P_{0,2} + \lambda_{cA_2}P_{1,1} \\
&\sum_{a=0}^{1}\sum_{b=1}^{2} P_{a,b} = 1
\end{aligned}\right\} \qquad (4.19)$$

Solving the above system of equations yields the probability $P_{a,b}$. Therefore, the probability, $Pb_{k,j}$, can be given by

$$Pb_{k,j} = \sum_{b=1}^{2} P_{1,b} \tag{4.20}$$

To determine the message waiting time, $Wc_{k,j}$, the network channel is treated as an MMPP/G/1 queueing system [31]. As the arrival process is modelled by MMPP_{cA} and the service time is $T_{k,j}$, $Wc_{k,j}$ can be obtained based on Eqs. (3.37) and (3.38).

Averaging over all possible nodes that can be destinations for a message in the ACN, the network latency for intra-cluster messages can be obtained as

$$T_{in} = \sum_{j=1}^{n} P_{j,\eta_{in}} T_j \tag{4.21}$$

where T_j is equal to the average service time of a message at stage 0, i.e., $T_j = T_{0,j}$.

4.3.2.2 Laplace-Stieltjes transform of the service time on network channels at stage *k*

Based on the determination of the message waiting time in Eqs. (3.37) and (3.38), the calculation of $Wc_{k,j}$ in Section 4.3.2.1 is driving the derivation of Laplace-Stieltjes transform of message service time on network channels in the ACN, $F^*_{k,j}(s)$. By virtue of Eq. (3.39), the Laplace-Stieltjes transform of $T_{k,j}$ can be expressed as

$$F^*_{k,j}(s) = \begin{cases} e^{-s\omega t^A_{cn}} & k = K-1 \\ W^*_{k,j}(s) e^{-s\omega t^A_{cs}} & 0 \le k \le K-2 \end{cases} \tag{4.22}$$

where $W^*_{k,j}(s)$ is the Laplace-Stieltjes transform of the blocking time experienced by

a $2j$-channel message on network channels at stage k, $Wb_{k,j}$. Examining the expression of $Wb_{k,j}$ in Eqs. (4.18)-(4.20), we can find that it is infeasible to find the exact expression for its Laplace transform. Based on Eqs. (3.43) and (3.45), the distribution of the message blocking time can be reasonably approximated by an exponential distribution and its Laplace-Stieltjes transform, $W^*_{k,j}(s)$, can be written as

$$W^*_{k,j}(s) = \frac{\alpha}{\alpha + s} \tag{4.23}$$

where α is selected to match the mean blocking time experienced by a message at network channels and can be found as

$$\alpha = \frac{1}{\sum_{\ell=k+1}^{K-1} Wb_{\ell,j}} \tag{4.24}$$

4.3.2.3 Waiting time at the source node

Recall that messages injected from the source node enter the ACN with the probability, $(1-\xi)$. Thus, the traffic arriving at an injection channel at the source node in the ACN, denoted by MMPP_{sA}, is a fraction of that generated by a source node. The fraction, f_{sA}, can be expressed as

$$f_{sA} = 1 - \xi \tag{4.25}$$

According to Eq. (3.18), the infinitesimal generator, $\mathbf{Q}_{sA}$, and rate matrix, $\mathbf{\Lambda}_{sA}$, of MMPP_{sA} can be easily obtained.

To determine the waiting time, W_{in}, that a message experiences before entering the network, the injection channel at the source is treated as an MMPP/G/1

queueing system, where the arrival process is modelled by MMPP_{sA} and the service time is the network latency for an intra-cluster message given by Eq. (4.21). W_{in} can be obtained based on Eqs. (3.37) and (3.38).

The time for the tail flit to reach the destination, τ_{in}, in intra-cluster communication networks can be simply found as [56]

$$\tau_{in} = \sum_{j=1}^{n} P_{j,\eta_{in}} \left(\sum_{k=1}^{K-1} t_{cs}^{A} + t_{cn}^{A} \right) \tag{4.26}$$

4.3.3 Message latency in inter-cluster communication networks

Inter-cluster messages traverse $2p$ channels in the ECN (p channels in the source cluster and p channels in the destination cluster) and $2q$ channels in the ICN to reach their destinations. Similarly to the intra-cluster messages, we take into account the network latency of a $2(p+q)$-channel message in inter-cluster communication networks (i.e., ECN and ICN) as T_{p+q}. The number of stages to be crossed for such a message is $K = 2(p+q)-1$. Based on Eqs. (4.14) and (4.16), the service time of a channel experienced by a message at stage k, $T_{k,p+q}$, in inter-cluster networks can be given by

$$T_{k,p+q} = \begin{cases} \sum_{\ell=k+1}^{K-1} Wb_{\ell,p+q} + \omega t_{cs}^{k} & 0 \le k \le K-2 \\ \omega t_{cn}^{E} & k = K-1 \end{cases} \tag{4.27}$$

where $Wb_{k,p+q}$ is the blocking time that a message acquires a channel at stage k in inter-cluster networks, and t_{cn}^{E} is the time for a flit to transmit on a node-to-switch

(or switch-to-node) connection in the ECN and can be obtained based on Eq. (4.15). $Wb_{k,p+q}$ can be calculated similarly to that used in the determination of $Wb_{k,j}$ in Eqs. (4.18)-(4.20). t_{cs}^{k} can be found according to the transmission time of each flit in the correspondence networks and can be expressed as

$$t_{cs}^{k} = \begin{cases} t_{cs}^{I} & p \le k < p + 2q - 1 \\ t_{cs}^{E} & \text{otherwise} \end{cases} \tag{4.28}$$

where t_{cs}^{E} and t_{cs}^{I} represent the time to transmit on a switch-to-switch connection in the ECN and the ICN, respectively, which can be determined by virtue of Eq. (4.17).

Similarly to intra-cluster communication networks, the network latency experienced by a message in inter-cluster communication networks, T_{out}, can be given by

$$T_{out} = \sum_{p=1}^{n} \sum_{q=1}^{n_c} P_{p,q} T_{p+q} \tag{4.29}$$

where T_{p+q} is equal to the service time of an inter-cluster message at stage 0, i.e., $T_{p+q} = T_{0,p,q}$; $P_{p,q}$ is the probability of a $2(p+q)$ -channel message and $P_{p,q} = P_p P_q$ (P_p and P_q can be computed by Eq.(4.6)).

Each inter-cluster message crosses a Concentrator (C) and a Dispatcher (D) during its network journey. The C/Ds are working as simple buffers and are used to combine traffic from/to one cluster to/from other clusters [17]. To determine the waiting time at these buffers, denoted by W_c and W_d, the C/D is modelled as an MMPP/G/1 queueing system. The arrival traffic injected into this queue is MMPP_{cI} and the service time is ωt_{cs}^{I}. To compute the waiting time experienced by a message

at the source in inter-cluster networks, W_{out}, the local queue is modelled as an MMPP/G/1 queueing system. The quantities, W_c, W_d, and W_{out}, can be determined based on Eqs. (3.37) and (3.38). Therefore, the latency experienced by a message in inter-cluster communication networks can be computed as follows:

$$Lat_{out} = T_{out} + W_{out} + W_c + W_d + \tau_{out} \tag{4.30}$$

where τ_{out} is the average time for the tail flit to reach the destination in inter-cluster communication networks and can be derived based on Eq. (4.26).

4.4 Validation of the model

A discrete-event simulator has been developed in the Objective Modular Network Testbed in C++ (OMNeT++) simulation environment [89], operating at the flit level, to validate the accuracy of the analytical model. The message latency is defined as the amount of time from the generation of a message until the last flit reaches the destination. Each simulation experiment was run until the network reached its steady state and collected based on Eqs. (3.51) and (3.52). The arrivals of messages generated by the source node are modelled by MMPP_s with infinitesimal generator $\mathbf{Q}_s$ and rate matrix $\mathbf{\Lambda}_s$. The message destinations are non-uniformly distributed over the network nodes according to the intra-cluster and inter-cluster communication locality.

Extensive simulation experiments have been performed to validate the accuracy of the analytical model for various combinations of cluster sizes, system sizes, switch sizes, message lengths, number of clusters, different MMPP_s traffic inputs, and various degrees of communication locality. However, for the sake of

Table 4.2: Network parameters for multi-cluster systems

Network	Bandwidth (per time unit)	Network Latency (time units)	Switch Latency (time units)
Net.1	800	0.02	0.01
Net.2	600	0.1	0.05

specific illustration and without loss of generality, latency results are presented for the 16 cluster system with the 4-port 3-tree to construct the ACN, ECN, and ICN. The system parameters are set as follows:

- Message length: $\omega = 64$ and 128 flits;
- Flit length: $L_f = 512$ and 1024 bytes;
- The configurations of each communication network in multi-cluster systems: ACN is configured by Net.1, ECN and ICN are configured by Net.2, where the configurations, i.e., Net.1 and Net.2, can be found in Table 4.2;
- The infinitesimal generator, $\mathbf{Q}_s$, of MMPP_s are set as follows, representing different degrees of traffic burstiness

$$\mathbf{Q}_s = \begin{bmatrix} -0.07 & 0.07 \\ 0.07 & -0.07 \end{bmatrix}, \quad \mathbf{Q}_s = \begin{bmatrix} -0.008 & 0.008 \\ 0.008 & -0.008 \end{bmatrix}, \quad \mathbf{Q}_s = \begin{bmatrix} -0.06 & 0.06 \\ 0.04 & -0.04 \end{bmatrix},$$

$$\mathbf{Q}_s = \begin{bmatrix} -0.09 & 0.09 \\ 0.03 & -0.03 \end{bmatrix}, \quad \mathbf{Q}_s = \begin{bmatrix} -0.03 & 0.03 \\ 0.015 & -0.015 \end{bmatrix}, \quad \mathbf{Q}_s = \begin{bmatrix} -0.08 & 0.08 \\ 0.06 & -0.06 \end{bmatrix},$$

$$\mathbf{Q}_s = \begin{bmatrix} -0.04 & 0.04 \\ 0.01 & -0.01 \end{bmatrix}, \text{ and } \mathbf{Q}_s = \begin{bmatrix} -0.07 & 0.07 \\ 0.035 & -0.035 \end{bmatrix};$$

- The intra-cluster locality (i.e., $\eta_{in} = 0.4, 0.6 \text{ and } 0.8$) and the inter-cluster locality (i.e., $\eta_{out} = \xi = 0.5, 0.7 \text{ and } 0.9$) are set in the caption of the figures, representing different degrees of communication locality.

Figs. 4.5 and 4.6 depict the performance results for the message latency predicted by the analytical model plotted against those provided by the simulator as a

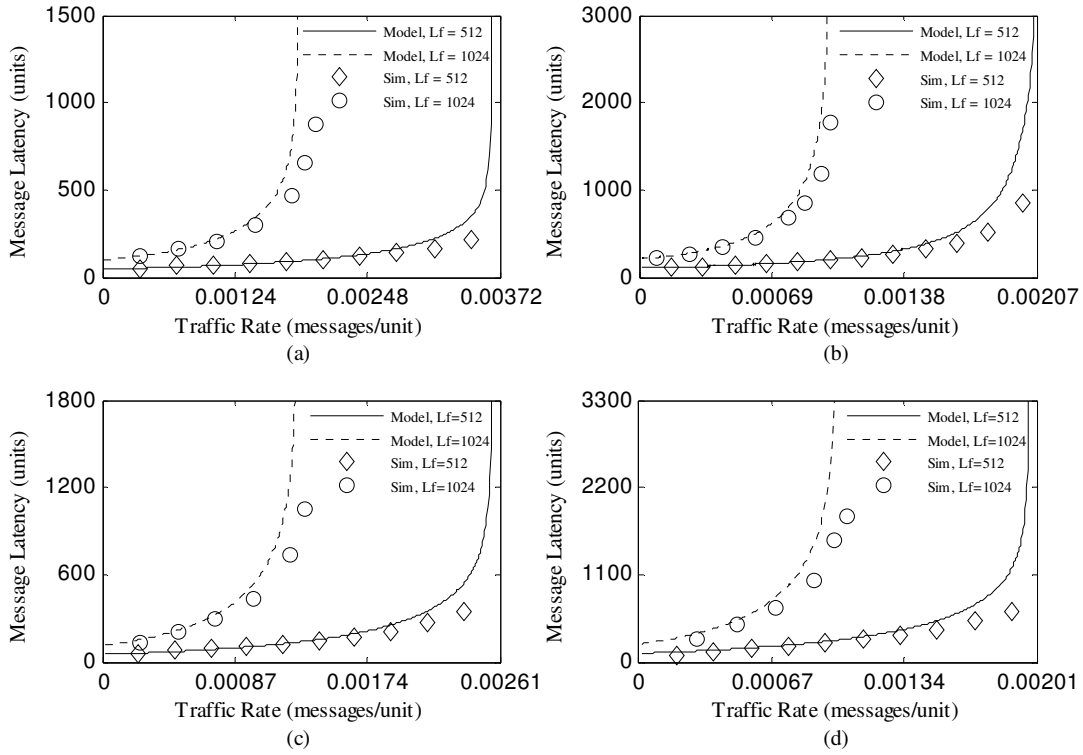

Figure 4.5: Latency predicted by the model and simulation: (a) $\eta_{in} = 0.6$, $\eta_{out} = 0.5$, $\omega = 64$, $\varphi_{s1} = 0.07$, $\varphi_{s2} = 0.07$, (b) $\eta_{in} = 0.6$, $\eta_{out} = 0.5$, $\omega = 128$, $\varphi_{s1} = 0.008$, $\varphi_{s2} = 0.008$, (c) $\eta_{in} = 0.8$, $\eta_{out} = 0.9$, $\omega = 64$, $\varphi_{s1} = 0.06$, $\varphi_{s2} = 0.04$, (d) $\eta_{in} = 0.8$, $\eta_{out} = 0.9$, $\omega = 128$, $\varphi_{s1} = 0.09$, $\varphi_{s2} = 0.03$

function of the traffic rate in multi-cluster systems. In these figures, the horizontal axis represents the traffic rate, λ_{s1}, at which a node injects messages into the network when the MMPP_s is at state 1, while the vertical axis denotes the mean message latency obtained from the model. For the sake of clarity of the figures, the arrival rate, λ_{s2}, at state 2 is deliberately set to zero; otherwise the three-dimensional graphs will be used to represent the results. These figures reveal that the message latency obtained from the analytical model closely match those obtained from the simulation. The tractability and accuracy of the analytical model make it a practical and cost-effective tool to gain insight into the performance of INs in multi-cluster systems in the presence of bursty traffic with communication locality.

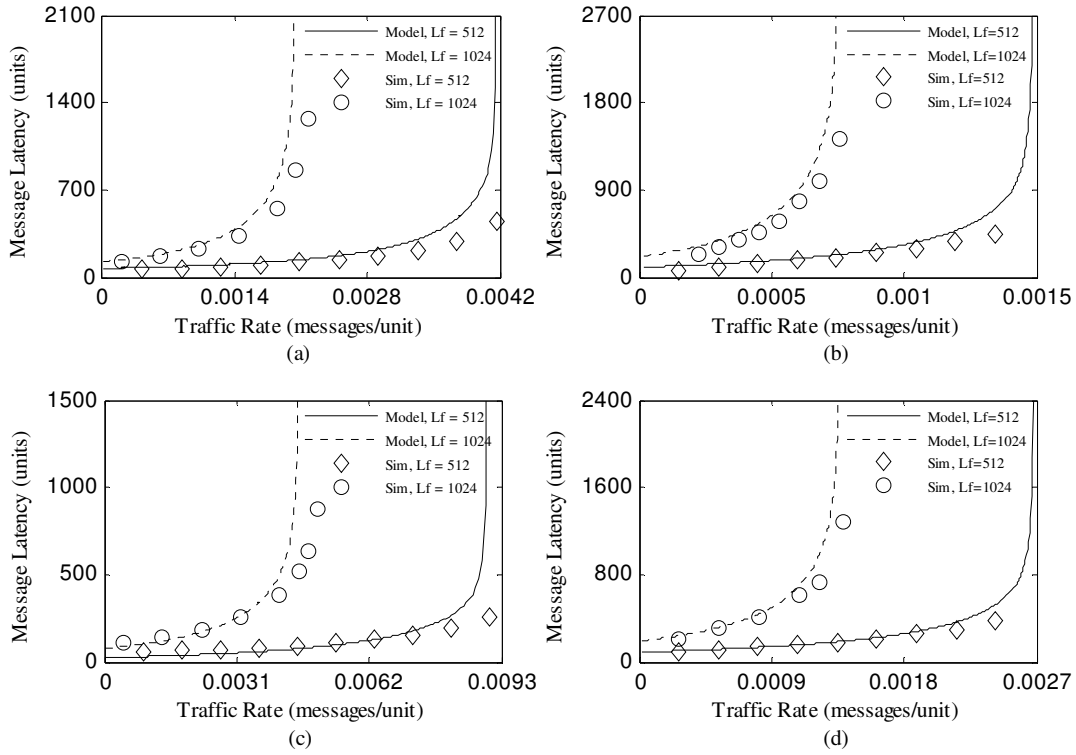

Figure 4.6: Latency predicted by the model and simulation: (a) $\eta_{in} = 0.8$, $\eta_{out} = 0.7$, $\omega = 64$, $\varphi_{s1} = 0.03$, $\varphi_{s2} = 0.015$, (b) $\eta_{in} = 0.8$, $\eta_{out} = 0.7$, $\omega = 128$, $\varphi_{s1} = 0.08$, $\varphi_{s2} = 0.06$, (c) $\eta_{in} = 0.4$, $\eta_{out} = 0.5$, $\omega = 64$, $\varphi_{s1} = 0.04$, $\varphi_{s2} = 0.01$, (d) $\eta_{in} = 0.4$, $\eta_{out} = 0.5$, $\omega = 128$, $\varphi_{s1} = 0.7$, $\varphi_{s2} = 0.35$

4.5 Performance analysis

Having validated the accuracy of the analytical model, in this section, the model will be used as a cost-effective tool to investigate the impact of bursty traffic with communication locality on the performance of INs in multi-cluster systems. Without loss of generality, we show a system with 32 clusters, and the ACN and ECN are constructed by an 8-port 3-tree topology and the ICN is constructed by an 8-port 2-tree topology. The message length is 32 flits and each flit contains 512 bytes. Each communication network in multi-cluster systems is configured as follows: ACN is configured by Net.1, ECN and ICN are configured by Net.2, where the configurations of Net.1 and Net.2 can be found in Table 4.2.

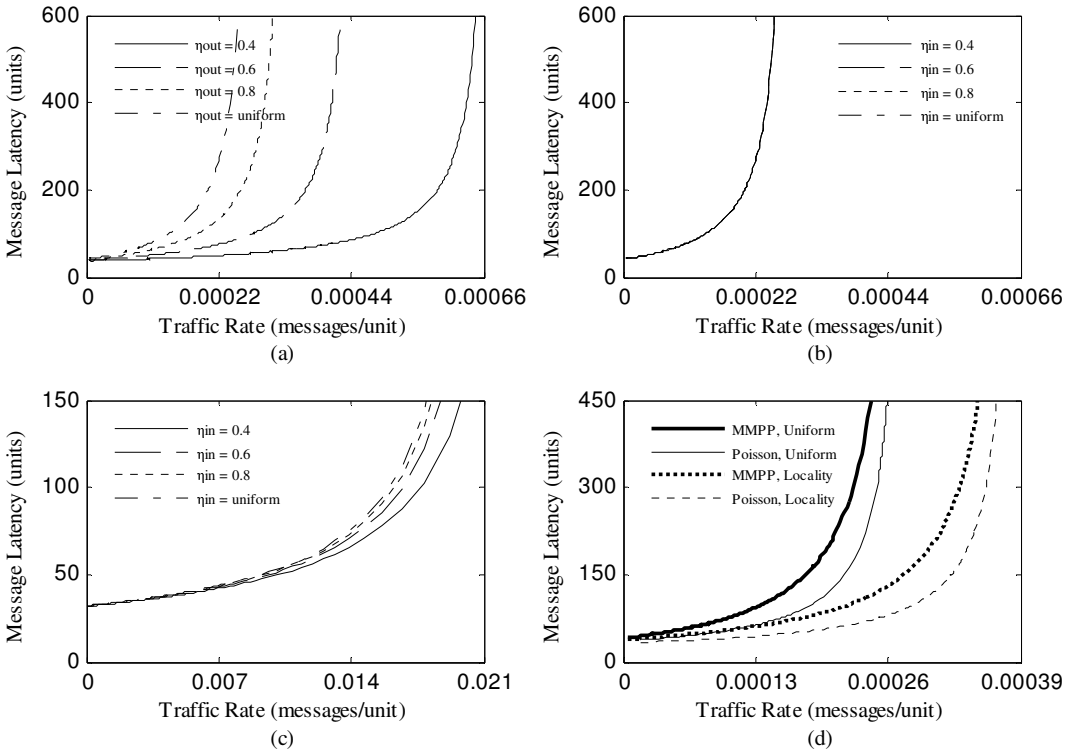

Figure 4.7: The impact of (a) inter-cluster locality, (b) (c) intra-cluster locality, and (d) bursty traffic on the performance of communication networks in multi-cluster systems

4.5.1 The impact of communication locality on network performance

In order to investigate the impact of communication locality on the performance of inter-cluster communication networks, Fig. 4.7(a) depicts the message latency when the inter-cluster locality, η_{out}, is set to 0.4, 0.6, 0.8, and $(C-1)N/(CN-1)$, where $\eta_{out} = (C-1)N/(CN-1)$ indicates the uniform destination distribution in inter-cluster communication networks, with the uniform destination distribution in the ACN. As can be seen from this figure, we can find that increasing the inter-cluster locality improves the overall system performance, since both the message latency decreases and the maximum network throughput goes up.

Similarly, in Fig. 4.7(b), we show the message latency in multi-cluster systems as the intra-cluster locality, η_{in}, varies from 0.4 to 1, where $\eta_{in} = 1$ represents the uniform destination distribution in intra-cluster communication networks, with the uniform destination distribution in inter-cluster communication networks. In this case, no significant impact appears on the overall system performance. To have a deeper understanding on the impact of intra-cluster locality, we investigate its effects on network performance in each individual cluster, as depicted in Fig. 4.7(c). As each individual cluster is homogeneous, the performance results shown in Fig. 4.7(c) can be obtained in any cluster. The same results appear as those shown in Fig. 4.7(a). We can conclude that the performance improvement in intra-cluster communication networks does not affect the overall system performance significantly, while the performance improvements in inter-cluster communication networks can directly affect the overall system performance.

4.5.2 The impact of bursty traffic on network performance

In order to quantitatively capture the impact of bursty traffic on system performance, in Fig. 4.7(d), we separately present the message latency when the network is subject to the bursty MMPP_s ($\varphi_{s1} = 0.6,\ \varphi_{s2} = 0.3$) against the non-bursty Poisson process (i.e., the assumption made in the existing study [54]) as a function of the mean arrival rate with and without the communication locality ($\eta_{in} = \eta_{out} = 0.7$), respectively. The results from the figure show that the model based on the Poisson assumption under-estimates the latency when the network is subject to the bursty traffic. Moreover, the bursty traffic can degrade the system performance seriously since both the message latency increases and the network saturates at a lower traffic rate. Our finding emphasises the importance of using realistic models to study and

optimise the performance of INs in multi-cluster systems. Furthermore, the use of communication locality can decrease the degrading effects of bursty traffic on the overall network performance in multi-cluster systems.

4.6 Conclusions

Driven by the importance and timeliness of evaluating and optimising system performance using the more realistic traffic loads, this chapter has presented a new analytical model to investigate the performance of INs in multi-cluster systems in the presence of bursty traffic with communication locality. Each communication network in the proposed system has been constructed as an m-port n-tree. Extensive simulation experiments have been conducted to validate the accuracy of the analytical model. The tractability and accuracy of the model make it a practical and cost-effective evaluation tool to gain insight into the performance of INs in multi-cluster systems in the presence of bursty traffic with communication locality. The results have shown that the performance improvement in intra-cluster communication networks does not affect the overall system performance significantly, while the performance improvements in inter-cluster communication networks can directly affect the overall system performance. The analytical results have also revealed that the bursty traffic can degrade the network performance considerably. Moreover, the results have demonstrated that the communication locality can decrease the degrading effects of bursty traffic on network performance in multi-cluster systems.

Chapter 5

An Analytical Model for Networks in Heterogeneous Multi-cluster Systems in the Presence of Bursty Traffic

Multi-cluster systems are embracing the environment with the even greater heterogeneity of computation and communication capacities in individual clusters [1, 115]. The underlying interconnection network (IN) is one of the key components that dominate the performance of heterogeneous multi-cluster systems [17]. Most existing analytical models on the performance study of INs have been limited to homogeneous cluster systems, and the evaluations are confined to a single cluster only [5, 16, 21, 41, 64, 74, 81, 92, 103, 122]. Many measurement studies [19, 110, 111] have demonstrated that the traffic generated by real-world applications exhibits a high degree of burstiness and possesses correlations in the number of message arrivals. To the best of our knowledge, there is only one analytical model reported in the current literature to investigate the performance of networks in multi-cluster systems taking bursty traffic into consideration. However, in this model, each cluster is modelled by an infinite-buffer queueing system without considering the specific network structure and components in each cluster, e.g., network topology, message switching and routing strategy, etc.

In order to fill this gap, this chapter presents a new analytical model to investigate the performance of INs in heterogeneous multi-cluster systems in the presence of bursty traffic. The model considers the specific architecture of the multi-cluster system, network topology, message switching and routing algorithm, etc. Moreover, this model is able to capture the heterogeneity in the number of nodes and the configuration of intra-communication networks within each cluster. In addition, a multi-user environment was taken into account in this chapter, i.e., the user level varies in different clusters. The accuracy of the model is demonstrated by comparing analytical results to those obtained through extensive simulation experiments.

The rest of this chapter is organised as follows. Section 5.1 shows the architecture of heterogeneous multi-cluster systems. The assumptions and notations adopted in this chapter are presented in Section 5.2. Section 5.3 derives the analytical model to investigate the network performance in heterogeneous multi-cluster systems. The results obtained from the analytical model are validated by extensive simulation experiments in Section 5.4. Section 5.5 carries out performance analysis. Finally, Section 5.6 concludes this study.

5.1 The architecture of heterogeneous multi-cluster systems

This chapter focuses on the heterogeneous multi-cluster system consisting of C clusters, as depicted in Fig. 5.1, where the i-th, $1 \le i \le C$, cluster contains N_i nodes with their own LMs. There are two communication networks in each cluster: an ACN used for the purpose of message passing between nodes within the same cluster and an ECN used to transmit messages between different clusters and manage the entire

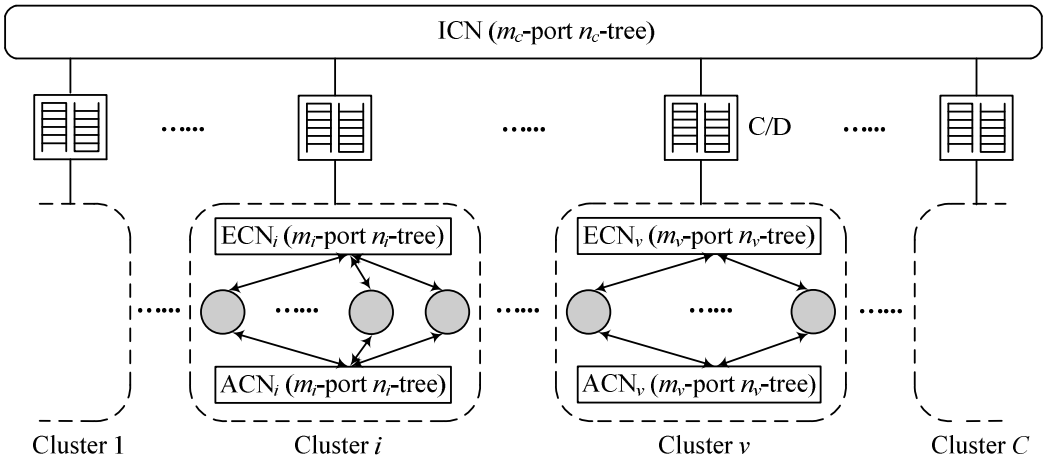

Figure 5.1: The architecture of heterogeneous multi-cluster systems

system [56, 120]. The clusters are connected to each other by an ICN. The ECN and ICN are connected by a set of C/Ds [17], which are used to combine message traffic from/to one cluster to/from other clusters. Each communication network (i.e., ACN, ECN, and ICN) in multi-cluster systems is constructed in the m-port n-tree. The specification of m-port n-tree can be found in Chapter 4.1.2. According to Eq. (4.2), the number of nodes, N_i, in the i-th cluster can be given by

$$N_i = 2\left(\frac{m_i}{2}\right)^{n_i} \tag{5.1}$$

To uniquely capture the heterogeneous number of nodes in each cluster, m_i and n_i are used to denote the parameters of the m-port n-tree topology in the i-th cluster. For clarity, the ACN_i and ECN_i with subscript i are adopted to represent the ACN and ECN in the i-th cluster.

The construction of ACN_i and ECN_i based on the m_i-port n_i-tree in the i-th cluster, is straightforward. In the ICN, each cluster can be viewed again as the node of the tree. Let m_c and (n_c+1) denote the switch size and the height of the tree

topology in the ICN. Since there are C clusters in the system, m_c and n_c should satisfy the following equation:

$$C = 2\left(\frac{m_c}{2}\right)^{n_c} \tag{5.2}$$

5.2 Assumptions and notations

The analytical model is based on the following assumptions, where Assumption (a) is used to capture the bursty nature of the realistic network traffic. The other assumptions are widely used in the related studies [5, 16, 21, 41, 56, 120]. A brief summary of key notations used in the derivation of the model is listed in Table 5.1.

a) The arrivals of messages generated by the source node are bursty and follow an MMPP_s (the subscript s denoting the traffic generated by the source node) with the infinitesimal generator, $\mathbf{Q}_s$, and rate matrix, $\mathbf{\Lambda}_s$. $\mathbf{Q}_s$ and $\mathbf{\Lambda}_s$ can be given by

$$\mathbf{Q}_s = \begin{bmatrix} -\varphi_{s1} & \varphi_{s1} \\ \varphi_{s2} & -\varphi_{s2} \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda}_s = \begin{bmatrix} \lambda_{s1} & 0 \\ 0 & \lambda_{s2} \end{bmatrix} \tag{5.3}$$

b) The message destinations are uniformly distributed over the network nodes;

c) Messages are routed through the network using the wormhole switching and deterministic routing, which have been illustrated in Chapter 4.1.3;

d) Message length is ω flits and the length of each flit is L_f bytes;

e) The configuration of the intra-cluster networks in each cluster and that of the inter-cluster networks are heterogeneous in terms of bandwidth, network latency and switch latency.

Table 5.1: Key notations used in the derivation of the model in Chapter 5

$\overline{Latency}$	mean message latency in the heterogeneous multi-cluster system
$\overline{Lat}_i$	message latency in heterogeneous multi-cluster systems from the perspective of cluster i
$P_{i,j}$	the probability that a newly generated message needs to cross $2j$ channels to reach its destination in the m_i-port n_i-tree
N_i	the number of nodes in cluster i
$\overline{d}_i$	average message distance in the m_i-port n_i-tree
ξ_i	the probability that a newly generated message in cluster i destines to the node in other clusters
C	the number of clusters in the system
ϕ_i	the parameter used to capture the multi-user level in cluster i
$\mathbf{Q}_s$, $\mathbf{\Lambda}_s$	parameter matrices of MMPP_s to model the traffic generated by the source node
$\mathbf{Q}_i^{cA}$, $\mathbf{\Lambda}_i^{cA}$	parameter matrices of MMPP_i^{cA} to model the traffic arriving at a given network channel in the ACN_i
MMPP_i^{in}	the traffic arriving at an injection channel to the ACN in cluster i
$\text{MMPP}_{i\to v}^{cE}$	the traffic arriving at a network channel in the ECN in cluster i to cluster v
MMPP^{cI}	the traffic arriving at a network channel in the ICN
$\overline{Lat}_i^{in}$	message latency experienced by a message in intra-cluster network in cluster i
$\overline{Lat}_i^{out}$	message latency experienced by a message in the inter-cluster networks from the perspective of cluster i
$\overline{W}_i^{in}$	waiting time at the source in the ACN_i
$W_{i\to v}^{out}$	waiting time that a message experiences at the source in inter-cluster networks from cluster i to cluster v
W_i^c, W_i^d	waiting time experienced by a message at the Concentrator and Dispatcher from the perspective of cluster i

$\overline{R}_i^{in}$	the time for the tail to reach the destination in intra-cluster networks in cluster i
$\overline{R}_{i\to v}^{out}$	the time for the tail to reach the destination in inter-cluster networks from cluster i to cluster v
T_i^j	the network latency of a $2j$-channel intra-cluster message in cluster i
$T_{i\to v}$	the network latency of a $(v+2r+u)$-channel inter-cluster message from cluster i to cluster v
$T_{i,k}^j$	the service time experienced by the $2j$-channel intra-cluster message on a channel at the stage k in cluster i
$T_{i\to v,k}$	the service time experienced by the $(v+2r+u)$-channel inter-cluster message on a channel at the stage k from cluster i to cluster v
$t_{cn_i}^A$, t_{cn}^E	the time for a single flit to transmit on a node-to-switch (or switch-to-node) connection in the ACN in cluster i and the ECN
$t_{cs_i}^A$, t_{cs}^E, t_{cs}^I	the time for a flit to transmit on a switch-to-switch connection in the ACN in cluster i, the ECN and the ICN
$Wb_{i,k}^j$	the blocking time that a $2j$-channel message acquires a channel at stage k in intra-cluster networks in cluster i
$Wb_{i\to v,k}$	the blocking time that a $(v+2r+u)$-channel message acquires a channel at stage k in inter-cluster networks from cluster i to cluster v
$Pb_{i,k}^j$	the probability that a $2j$-channel intra-cluster message is blocked at stage k in cluster i
$Wc_{i,k}^j$	the waiting time experienced by a $2j$-channel intra-cluster message to acquire a channel when blocking occurs at stage k in cluster i

5.3 The analytical model

Based on Eq. (4.6), the probability, $P_{i,j}$, that a newly generated message need cross $2j$ channels (j channels in ascending phase and j channels in descending phase) to reach its destination in the m_i-port n_i-tree is given by

$$P_{i,j}=\begin{cases}\dfrac{(m_i-1)\left(\dfrac{m}{2}\right)^{j-1}}{N_i-1} & j=n_i \\ \dfrac{\left(\dfrac{m_i}{2}-1\right)\left(\dfrac{m}{2}\right)^{j-1}}{N_i-1} & 1\le j\le n_i-1\end{cases} \tag{5.4}$$

where N_i is the number of nodes in cluster i and can be given by Eq. (5.1). As a result, the average message distance, $\bar{d}_i$, in the m_i-port n_i-tree can be computed by

$$\bar{d}_i=\sum_{j=1}^{n_i}2jP_{i,j} \tag{5.5}$$

In what follows, the analytical model will be derived to investigate the performance of INs in heterogeneous multi-cluster systems. To this end, the traffic characteristics on network channels will be first investigated in Chapter 5.3.1 and then the message latency in intra-cluster and inter-cluster communication networks will be determined in Chapters 5.3.2 and 5.3.3, respectively. Finally, the message latency of the whole system will be obtained in Chapter 5.3.4.

5.3.1 Traffic characteristics on channels in intra-cluster and inter-cluster communication networks

With the uniform distribution of message destinations, the probability, ξ_i, that a newly generated message in cluster i is destined to a node in other clusters can be determined as follows:

$$\xi_i=\frac{\sum_{k=1,k\neq i}^{C}N_k}{\sum_{k=1}^{C}N_k-1} \tag{5.6}$$

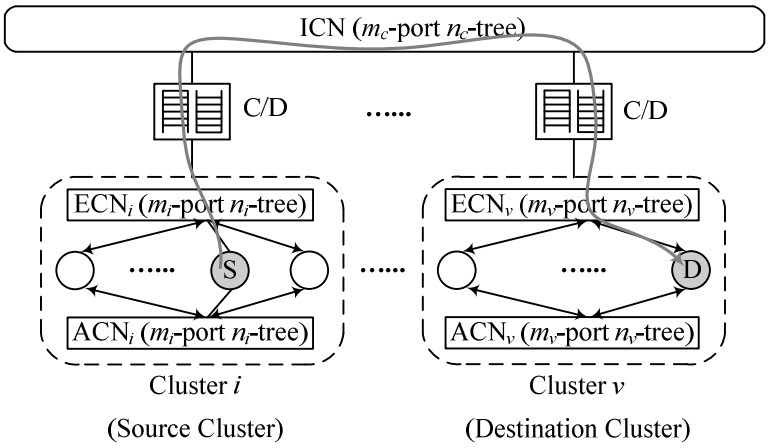

Figure 5.2: The message path in inter-cluster communication networks

where C is the number of clusters in the system and N_k is the number of nodes in the k-th cluster. Let us refer to the message with the source and destination in the same cluster as *intra-cluster messages*. Other messages are referred to as *inter-cluster messages*.

This model considers the multi-user environment in multi-cluster systems. To capture the characteristics of this type of environment, the multi-user level is assumed to be ϕ_i, where $0 \le \phi_i < 1$, for all users (i.e., nodes) in cluster i. This multi-user level implies the overhead since the system is in the multi-user environment [20], therefore, the effective traffic sent out by a node in cluster i can be approximated as $(1-\phi_i)^{-1}$ times the traffic generated by the node. Due to the traffic uniformity in intra-cluster networks, the message arrival process exhibits similar statistical behaviour over all network channels. Since an intra-cluster message in cluster i needs to cross, on average, d_i channels to reach its destination, the traffic arriving at a network channel in intra-cluster communication networks (i.e., ACN_i) is f_i^A times of the traffic generated by a source node. Since the total number of channels in the ACN_i is $4n_iN_i$, f_i^A can be written as

$$f_i^A = \frac{N_i \overline{d}_i (1-\xi_i)}{4 n_i N_i} (1-\phi_i)^{-1} = \frac{\overline{d}_i (1-\xi_i)}{4 n_i} (1-\phi_i)^{-1} \tag{5.7}$$

Recall that, in inter-cluster communication networks, messages generated from the source nodes in cluster i have the probability, ξ_i, to enter the ECN_i. The inter-cluster message leaves the ECN_i, crosses through the ICN and then is injected to the ECN_v in cluster v to reach its destination node, as shown in Fig. 5.2. Since the m_i-port n_i-tree is not a node-symmetric topology [17] and the size of each cluster may be different, a simple and efficient way is adopted to handle this asymmetric problem in inter-cluster networks in order to determine the traffic characteristics on network channels in the ECN and ICN, i.e., to identify the traffic loads from the perspective of each cluster and then average over all clusters. Based on Eq. (5.7), the traffic arriving at a given network channel in the ECN in cluster i to cluster v, is $f_{i \to v}^E$ times of the traffic generated by a source node. Taking the number of channels in the ECN from the perspective of cluster i into account, $f_{i \to v}^E$ can be given by

$$f_{i \to v}^E = \frac{[N_i \xi_i (1-\phi_i)^{-1} + N_v \xi_v (1-\phi_v)^{-1}] \overline{d}_i}{4 n_i N_i} \tag{5.8}$$

Similarly, the traffic arriving at a network channel in the ICN is f^I times of the traffic generated by a source node. Therefore,

$$f^I = \frac{\sum_{i=1}^{C} N_i \xi_i (1-\phi_i)^{-1} \overline{d}_c}{4 n_c C} \tag{5.9}$$

where $\overline{d}_c$ is the average number of channels that a message traverses to reach its destination in the ICN (i.e., m_c-port n_c-tree) and can be determined by Eq. (5.5).

Given that the splitting and superposition of multiple MMPPs give rise to a new MMPP [31], let MMPP_i^{cA}, $\text{MMPP}_{i\to v}^{cE}$, and MMPP^{cI} denote the traffic arriving at a given network channel in the ACN_i, in the ECN in cluster i to cluster v, and in the ICN, respectively. By virtue of Eqs. (3.17)-(3.34), the parameter matrices of MMPP_i^{cA}, $\text{MMPP}_{i\to v}^{cE}$ and MMPP^{cI} can be easily determined.

5.3.2 Message latency in intra-cluster communication networks

For the sake of illustration, the switches in the network are numbered based on the method presented in Chapter 4.3.2.1, i.e., the numbering starts from the switch next to the source node (STAGE 0) and goes up as we get closer to the destination node (STAGE $K-1$). Based on the *m*-port *n*-tree topology shown in Fig. 4.1, there are two types of connections: node-to-switch (or switch-to-node) connection and switch-to-switch connection.

5.3.2.1 Network latency for intra-cluster messages

Since an intra-cluster message may cross different numbers of channels to reach its destination, we take into account the network latency of a $2j$-channel message (i.e., a message need traverse $2j$ channels to reach its destination) as T_i^j, where the subscript i denotes the calculation from the perspective of cluster i. In m_i-port n_i-tree, the number of stages to be crossed for a $2j$-channel message is $K = 2j-1$.

Since messages are transferred to the local nodes as soon as they arrive at their destinations, we first consider the service time experienced by a message at the final stage and then continue the analysis backward to the first stage. Based on Eq.

(4.14), the service time experienced by a $2j$-channel message at stage $K-1$, $T_{i,K-1}^{j}$, can be given by

$$T_{i,K-1}^{j} = \omega t_{cn_i}^{A} \tag{5.10}$$

where $t_{cn_i}^{A}$ is the time to transmit from node-to-switch (or switch-to-node) connection in the ACN in cluster i and can be determined by Eq. (4.15).

The service time, $T_{i,k}^{j}$, experienced by the $2j$-channel message at the internal stages k can be computed by the actual message transmission time and the delay due to blocking at subsequent stages. Based on Eq. (4.16), $T_{i,k}^{j}$ can be given by

$$T_{i,k}^{j} = \omega t_{cs_i}^{A} + \sum_{\ell=k+1}^{K-1} Wb_{i,\ell}^{j}, \quad 0 \le k \le K-2 \tag{5.11}$$

where $t_{cs_i}^{A}$ is the time to transmit from switch-to-switch connection in the ACN in cluster i and can be calculated by Eq. (4.17). $Wb_{i,k}^{j}$ is the blocking time that a $2j$-channel message experiences to acquire a channel at stage k in cluster i. Based on Eq. (4.18), $Wb_{i,k}^{j}$ can be expressed as

$$Wb_{i,k}^{j} = Pb_{i,k}^{j} Wc_{i,k}^{j} \tag{5.12}$$

where $Pb_{i,k}^{j}$ is blocking probability experienced by a $2j$-channel message at stage k in cluster i and $Wc_{i,k}^{j}$ is the waiting time experienced by the message to acquire a channel when blocking occurs. The blocking probability, $Pb_{i,k}^{j}$, can be determined based on Eqs. (4.19) and (4.20). To determine the waiting time, $Wc_{i,k}^{j}$, a network channel is treated as an MMPP/G/1 queueing system [31]. Following Eqs. (3.37) and

(3.38), the waiting time, $Wc_{k,j}^{i}$, can be readily obtained.

Averaging over all the possible nodes destined made by a message, the network latency in intra-cluster networks can be obtained as

$$\overline{T}_i^{in} = \sum_{j=1}^{n_i} P_{i,j} T_i^j \tag{5.13}$$

where T_i^j is equal to the service time of a message at stage 0, i.e., $T_i^j = T_{i,0}^j$.

5.3.2.2 Waiting time at the source node

Recall that messages generated by a source node in cluster i enter the ACN with the probability, $(1-\xi_i)$. The traffic arriving at an injection channel to the ACN in cluster i, denoted by MMPP_i^{in}, is f_i^{in} times of that generated by a source node. Taking the overhead due to the multi-user environment into account, f_i^{in} can be expressed as

$$f_i^{in} = (1-\xi_i)(1-\phi_i)^{-1} \tag{5.14}$$

The parameter matrices of MMPP_i^{in} can be obtained based on Eqs. (3.17)-(3.34). To determine the waiting time, $\overline{W}_i^{in}$, that a message experiences at the source node in intra-cluster networks, the local queue is modelled by an MMPP/G/1 queueing system with the arrival process, MMPP_i^{in}, and the service time, $\overline{T}_i^{in}$. The waiting time, $\overline{W}_i^{in}$, can be determined by virtue of Eqs. (3.37) and (3.38).

Based on Eq. (4.13), the message latency, $\overline{Lat}_i^{in}$, in intra-cluster communication network in cluster i can be expressed as

$$\overline{Lat}_i^{in} = \overline{T}_i^{in} + \overline{W}_i^{in} + \overline{R}_i^{in} \tag{5.15}$$

where $\overline{R}_i^{in}$ is the time for the tail to reach the destination in intra-cluster networks and can be obtained based on Eq. (4.26) as follows:

$$\overline{R}_i^{in} = \sum_{j=1}^{n_i} P_{i,j}\left(\sum_{k=1}^{K-1} t_{cs_i}^A + t_{cn_i}^A\right) \tag{5.16}$$

5.3.3 Message latency in inter-cluster communication networks

As can be seen from Fig. 5.2, an inter-cluster message traverses both the ECN and the ICN to reach its destination. Specifically, each inter-cluster message crosses υ channels in the ECN_i in the source cluster i, $2r$ channels in the ICN, and u channels in the ECN_v in the destination cluster v to reach its destination node. Therefore, the number of stages for an inter-cluster message is $K = \upsilon + 2r + u - 1$.

The service time of a channel in inter-cluster networks from the perspective of cluster i can be obtained similarly to that in the intra-cluster networks. Thus, based on Eqs. (5.10) and (5.11), the service time, $T_{i\to v,k}$, experienced by a message on the channels at stage k in inter-cluster networks from the perspective of cluster i can be obtained as

$$T_{i\to v,k} = \begin{cases} \omega t_{cn}^E & k = K-1 \\ \omega t_{cs}^k + \sum_{\ell=k+1}^{K-1} Wb_{i\to v,\ell} & 0 \le k \le K-2 \end{cases} \tag{5.17}$$

where $Wb_{i\to v,k}$ is the blocking time that a message experiences to acquire a channel at stage k in inter-cluster networks and can be obtained according to Eq. (5.12). t_{cs}^k can be found according to the time to transmit each flit in the correspondence

networks. Based on Eq. (4.28), t_{cs}^{k} can be expressed as

$$t_{cs}^{k} = \begin{cases} t_{cs}^{I} & \upsilon \le k < \upsilon + 2r - 1 \\ t_{cs}^{E} & otherwise \end{cases} \tag{5.18}$$

where t_{cn}^{E} and t_{cs}^{E} represent the time to transmit from node-to-switch (or switch-to-node) connection and switch-to-switch connection in the ECN, and t_{cs}^{I} corresponds to the time to transmit from switch-to-switch connection in the ICN. These quantities can be obtained based on Eqs. (4.15) and (4.17). The network latency experienced by a message in inter-cluster networks can be calculated as follows:

$$T_{i\to v}^{out} = \sum_{\upsilon=1}^{n_i} \sum_{r=1}^{n_c} \sum_{u=1}^{n_v} P^{out} T_{i\to v} \tag{5.19}$$

where $T_{i\to v}$ is the latency for a $(\upsilon + 2r + u)$ -channel inter-cluster message, and $T_{i\to v} = T_{i\to v,0}$; P^{out} is the probability that a message needs to cross $(\upsilon + 2r + u)$ channels to reach its destination and can be given by $P^{out} = P_{i,\upsilon} P_{c,r} P_{v,u}$.

To determine the waiting time, $\overline{W}_{i\to v}^{out}$, that a message experiences at the source node in inter-cluster networks, the local queue is modelled by an MMPP/G/1 queueing system. $\overline{W}_{i\to v}^{out}$ can be obtained based on Eqs. (3.37) and (3.38).

The message latency, $\overline{Lat}_{i}^{out}$, in the inter-cluster networks can be determined by averaging latencies for the message from cluster *i* to all the other clusters, and can be expressed as

$$\overline{Lat}_{i}^{out} = \frac{\sum_{v=1, v\neq i}^{C} \left(\overline{T}_{i\to v}^{out} + \overline{W}_{i\to v}^{out} + \overline{R}_{i\to v}^{out} \right)}{C-1} \tag{5.20}$$

where $\overline{R}_{i\to v}^{out}$ is the time for the tail flit to reach the destination in inter-cluster networks and can be obtained according to Eq. (5.16).

5.3.4 Mean message latency in heterogeneous multi-cluster systems

The inter-cluster messages need traverse a Concentrator and a Dispatcher during its network journey, as shown in Fig. 5.2. The C/Ds are used to combine traffic from/to one cluster to/from other clusters and are working as simple buffers [17]. The waiting time experienced by a message at the Concentrator and Dispatcher from the perspective of cluster i, denoted by W_i^c and W_i^d, can be determined based on Eqs. (3.37) and (3.38).

The message latency in heterogeneous multi-cluster systems from the perspective of cluster i can be computed by

$$\overline{Lat}_i = (1-\xi_i)\overline{Lat}_i^{in} + \xi_i(\overline{Lat}_i^{out} + W_i^c + W_i^d) \tag{5.21}$$

The mean message latency in the whole system can be determined using a weighted arithmetic average as follows:

$$\overline{Latency} = \sum_{i=1}^{C}\left(\frac{N_i}{\sum_{j=1}^{C} N_j}\overline{Lat}_i\right) \tag{5.22}$$

5.4 Validation of the model

A discrete-event simulator has been developed based on the OMNeT++ simulation environment [89], operating at the flit level, to validate the accuracy of the analytical

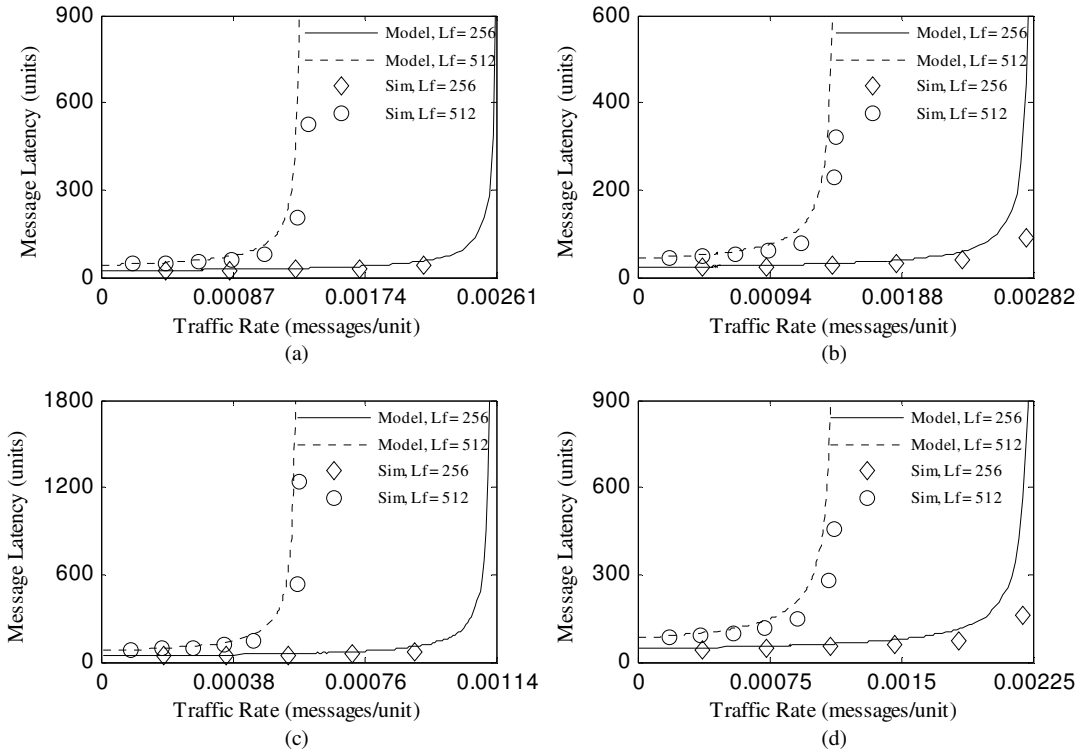

Figure 5.3: Latency predicted by the model and simulation in 8 cluster systems with (a) $\omega = 32$, $\varphi_{s1} = 0.08$, $\varphi_{s2} = 0.06$, (b) $\omega = 32$, $\varphi_{s1} = 0.06$, $\varphi_{s2} = 0.04$, (c) $\omega = 64$, $\varphi_{s1} = 0.007$, $\varphi_{s2} = 0.007$, and (d) $\omega = 64$, $\varphi_{s1} = 0.09$, $\varphi_{s2} = 0.03$

model. Each simulation experiment was run until the network reached its steady state and collected based on Eqs. (3.51) and (3.52). The arrivals of messages generated by the source node are modelled by MMPP$_s$ with infinitesimal generator $\mathbf{Q}_s$ and rate matrix $\mathbf{\Lambda}_s$.

Extensive simulation experiments have been performed to validate the accuracy of the analytical model. However, for the sake of specific illustration, Figs. 5.3 and 5.4 depict the analytical results for the message latency predicted by the analytical model plotted against those obtained by the simulator as a function of the traffic rate in heterogeneous multi-cluster systems with the following cases:

- Message length is $\omega = 32$ and 64 flits;
- Flit length is $L_f = 256$ and 512 bytes;

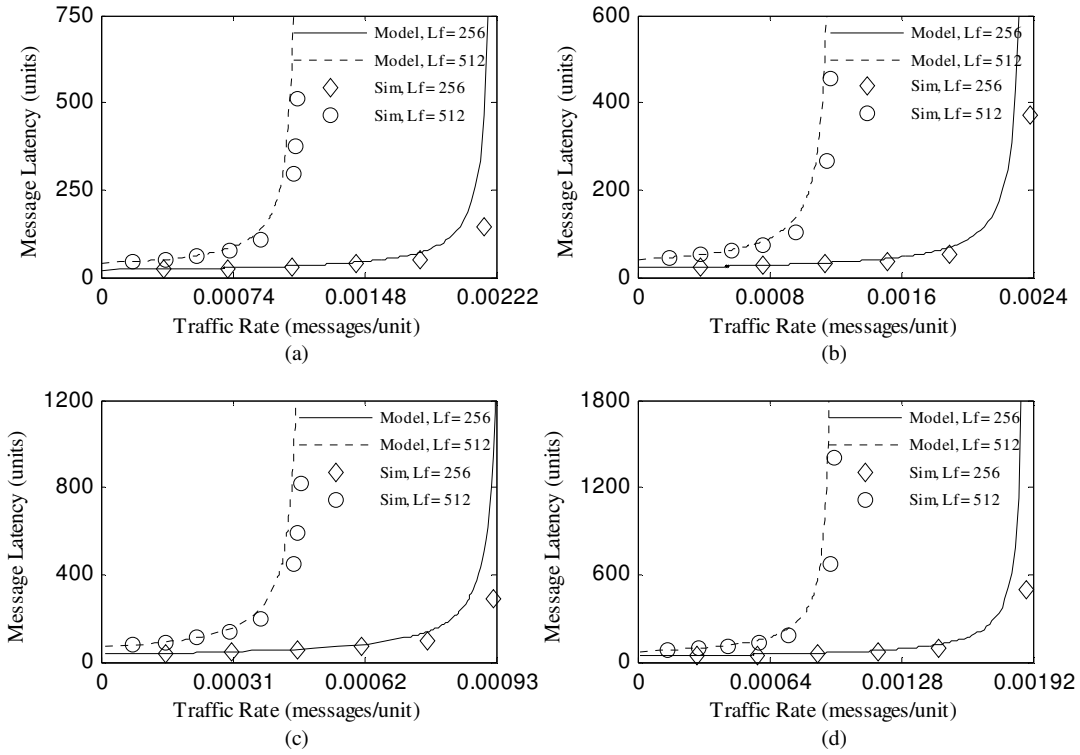

Figure 5.4: Latency predicted by the model and simulation in 16 cluster systems with (a) $\omega = 32$, $\varphi_{s1} = 0.08$, $\varphi_{s2} = 0.06$, (b) $\omega = 32$, $\varphi_{s1} = 0.06$, $\varphi_{s2} = 0.04$, (c) $\omega = 64$, $\varphi_{s1} = 0.007$, $\varphi_{s2} = 0.007$, and (d) $\omega = 64$, $\varphi_{s1} = 0.09$, $\varphi_{s2} = 0.03$

- The network parameters and system configurations can be found in Tables 5.2 and 5.3, respectively. In Table 5.3, N is the number of nodes in the whole system, C is the number of clusters, m_i-port n_i-tree is the network topology used to construct the ACN and ECN in cluster i, ϕ_i is used to indicate the user level in cluster i, and m_c-port n_c-tree is used for the construction of the ICN;
- The infinitesimal generator, $\mathbf{Q}_s$, of MMPP$_s$ is set in the captions of the figures, representing different degrees of traffic burstiness and correlations.

In these figures, the horizontal axis represents the traffic rate, λ_{s1}, at which a node injects messages into the network when the MMPP$_s$ is at state 1, while the vertical axis denotes the message latency obtained from the analytical model. For the

Table 5.2: Network parameters for heterogeneous multi-cluster systems

Network	Bandwidth (per time unit)	Network Latency (time units)	Switch Latency (time units)
Net.1	1000	0.01	0.005
Net.2	800	0.01	0.01
Net.3	600	0.04	0.06
Net.4	400	0.1	0.05
Net.5	200	0.2	0.1

Table 5.3: System configuration parameters for heterogeneous multi-cluster systems

N	*C*	Cluster organizations
376	8	$n_i = 1,\ \phi_i = 0.1,\quad i \in [0,2]$, ACN = **Net.3** $n_i = 2,\ \phi_i = 0.15,\ i \in [3,5]$, ACN = **Net.4** $n_i = 3,\ \phi_i = 0.2,\quad i \in [6,7]$, ACN = **Net.5** $m_i = m_c = 8\ \ (1 \le i \le C)$ ECN = **Net.2**, ICN = **Net.1**
800	16	$n_i = 2,\ \phi_i = 0.05,\ i \in [0,7]$, ACN = **Net.3** $n_i = 4,\ \phi_i = 0.1,\quad i \in [8,10]$, ACN = **Net.4** $n_i = 6,\ \phi_i = 0.15,\ i \in [11,15]$, ACN = **Net.5** $m_i = m_c = 4\ \ (1 \le i \le C)$ ECN = **Net.2**, ICN = **Net.1**

sake of clarity of the figures, the arrival rate, λ_{s2}, at state 2 is deliberately set to zero; otherwise the three-dimensional graphs will be used to represent the results. The figure reveals that the analytical results of the message latency closely match those obtained from the simulation. The tractability and accuracy of the analytical model make it a practical and cost-effective tool to gain insight into the performance of INs in heterogeneous multi-cluster systems in the presence of bursty traffic.

5.5 Performance analysis

Having validated the accuracy of the analytical model, in this section, the model is used as a cost-effective tool to investigate the trade-off between the size of each

cluster and the number of clusters in multi-cluster systems with the aim of maximising the network performance under bursty traffic. Without loss of generality, we consider a multi-cluster system with 2048 nodes to be constructed as follows:

1) 4 clusters with 512 nodes in each cluster (i.e., $C = 4$ and $N_i = 512$);
2) 8 clusters with 256 nodes in each cluster (i.e., $C = 8$ and $N_i = 256$);
3) 16 clusters with 128 nodes in each cluster (i.e., $C = 16$ and $N_i = 128$);
4) 32 clusters with 64 nodes in each cluster (i.e., $C = 32$ and $N_i = 64$).

To make these scenarios comparable, the bandwidth of each system is further normalised to make sure the total bandwidth of the whole system in each scenario is the same. To achieve this, the total bandwidth of the system in the first scenario is used as the benchmark and then the bandwidth in the other scenarios is adjusted. Specifically, let $\hat{B}$ denote the total bandwidth of the whole system and B_i^A, B_i^E and B^I represent the bandwidth of one network channel in the ACN_i, ECN_i and ICN, respectively. $\hat{B}$ can be expressed as

$$\hat{B} = \sum_{i=1}^{C} 4n_i N_i B_i^A + \sum_{i=1}^{C} 4n_i N_i B_i^E + 4n_c C B^I \tag{5.23}$$

For the sake of illustration, $B_i^A = B_i^E = B^I = 1000$ is set in the first scenario. Based on Eq. (5.23), B_i^A, B_i^E and B^I in the other scenarios can be obtained.

In order to quantitatively evaluate the impact of bursty nature of message arrivals on network performance, Fig. 5.5 depicts the message latency in the above scenarios when the network is subject to the bursty MMPP_s and non-bursty Poisson process as a function of the mean arrival rate. The mean arrival rate of MMPP_s can be determined according to Eq. (2.11). In this figure, the infinitesimal generator of

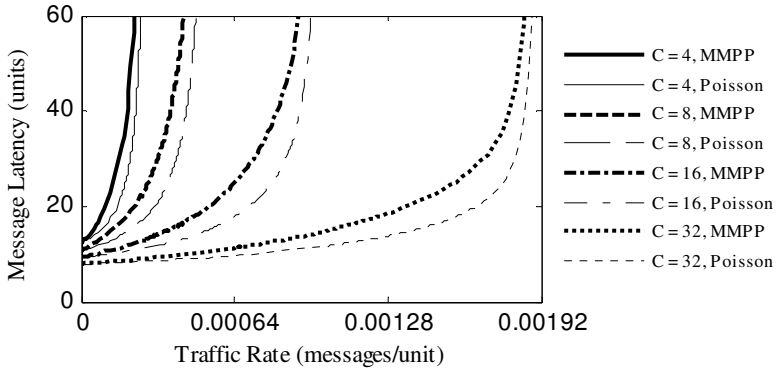

Figure 5.5: Latency predicted by the model with $\omega = 32$ flits and $L_f = 256$ bytes in the scenarios (1)-(4)

MMPP_s is set to $\varphi_{s1} = 0.09$ and $\varphi_{s2} = 0.09$. The degree of traffic burstiness and correlations can be obtained based on Eqs. (2.14) and (2.15). We can find from the figure that the model based on the non-bursty Poisson assumption seriously under-estimates the latency when the network is subject to the bursty traffic, especially under moderate and heavy traffic loads. The result is consistent with that is concluded in Chapter 4.5.2.

The performance results shown in Fig. 5.5 also reveal that increasing the number of clusters can improve the network performance since the message latency decreases and the maximum network throughput increases. Furthermore, the improvement caused by the trade-off between the size of each cluster and the number of overall clusters in the system can relieve the negative effects of bursty traffic on network performance. This observation is important since we can use existing economical network architectures in multi-cluster systems, which cannot ensure efficient support for bursty traffic in single-cluster systems, by increasing the number of clusters. However, the space occupied by the system becomes larger with an increasing number of clusters. Therefore, it is critical to balance the size of each cluster and the number of overall clusters within a limited space to maximise the

network performance.

5.6 Conclusions

The heterogeneous properties of individual clusters in multi-cluster systems can significantly affect the performance of their underlying INs. This chapter has developed a new analytical model which can be used to investigate the performance of INs in heterogeneous multi-cluster systems. The model has been able to characterise the heterogeneity in the number of nodes and the configuration of intra-communication networks within each cluster. Moreover, the model has considered a multi-user environment in multi-cluster systems. In addition, to capture the traffic arrival patterns of realistic network applications, the model has been derived in the presence of bursty traffic. The comparison between the results from the analytical model and those obtained from extensive simulation experiments has shown that the derived analytical model possesses a good degree of accuracy for predicting the network performance under different design alternatives and various traffic conditions. The results of performance analysis have highlighted the importance of using the realistic traffic models in order to quantitatively capture the performance properties of INs in heterogeneous multi-cluster systems. Moreover, the results have revealed the importance of balancing the size of each cluster and the number of clusters in order to maximise the overall system performance.

Chapter 6

A Performance Model for Integrated Wireless Networks in the Presence of Bursty Traffic with Communication Locality

The increasing demand for the coverage of WLANs is driving the installation of a very large number of APs [6, 25]. The traditional deployment of APs requires wired connection between APs and thus introduces high complexity and costs for their deployment in certain areas [114]. WMNs have emerged as a promising technology in next generation networks to provide economical and scalable broadband access to wireless interconnection of multiple APs which manage individual WLANs in order to extend the coverage of conventional single WLANs [4, 39, 46, 57]. WMNs are constructed by MRs, which form a multi-hop ad-hoc network [4, 123]. The stations in different WLANs are able to communicate with each other through relaying packets among MRs in a multi-hop fashion. A number of measurement studies [10, 49, 86] have demonstrated that the traffic generated by many real-world applications in wireless networks exhibits a high degree of burstiness and the message destinations are non-uniformly distributed over the network nodes.

Analytical models for WLANs and WMNs have been widely reported in the

literature [7, 8, 11, 33, 34, 50, 51, 67, 71, 76]. However, all these models have focused on the performance study of WLAN or WMN only. There has not been any analytical model reported in the current literature to investigate the QoS performance measures in the integrated WLANs and WMN when the network is subject to bursty traffic with non-uniformly distributed message destinations. To this end, this chapter presents an integrated wireless network where the WMN is used to connect multiple APs managing individual WLANs. In particular, a new analytical model is proposed to evaluate the end-to-end delay in the WMN interconnecting multiple WLANs in the presence of bursty traffic. Moreover, the proposed model can capture the phenomenon of communication locality. The specification of communication locality can be found in Section 4.1.4. The accuracy of the analytical model is validated through extensive simulation experiments. The derived analytical model is then used as a cost-effective tool to obtain the maximum achievable throughput which reflects the capacity of wireless networks. Moreover, the analytical model is used to investigate the impact of bursty traffic with communication locality on network performance, i.e., end-to-end delay and maximum achievable throughput.

The rest of this chapter is organised as follows. Section 6.1 shows useful preliminaries including the architecture of integrated WLANs and WMN, the network topology, and the routing strategy and MAC scheme in integrated wireless networks. Section 6.2 derives the analytical model to investigate the network performance. The results obtained from the model are validated in Section 6.3. Section 6.4 carries out performance analysis. Finally, Section 6.5 concludes this study.

6.1 Preliminaries

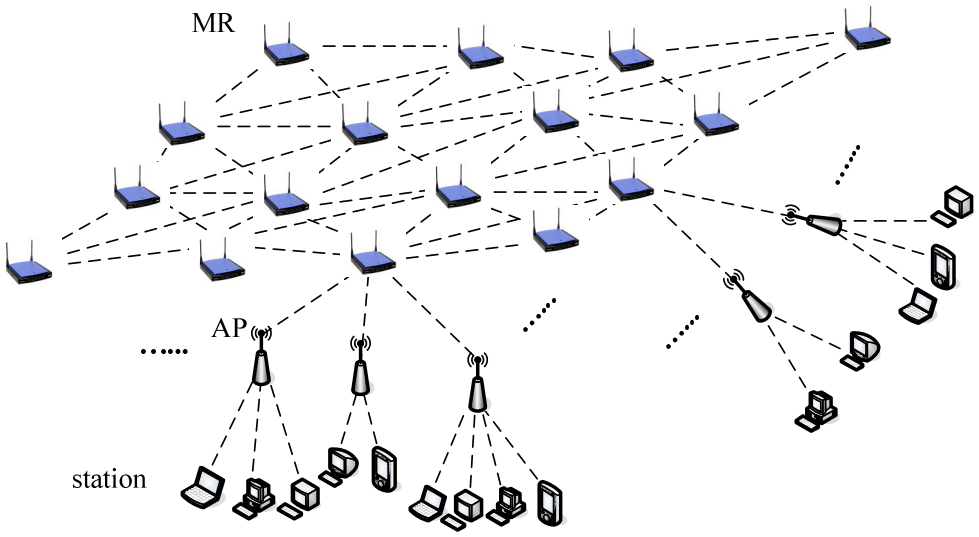

Figure 6.1: The architecture of integrated WLANs and WMN

6.1.1 The architecture of integrated WLANs and mesh networks

This study focuses on the WMNs interconnecting multiple WLANs [25, 119], as depicted in Fig. 6.1, where the MRs form the multi-hop mesh network. One AP manages an individual WLAN. The dashed lines indicate the possible communications between each pair of entities. Since a moderate level of perturbations from ideal grid placement is suitable for the performance study of WMNs to be illustrated in Section 6.1.2, we consider the situation that C MRs are uniformly distributed in an $n \times n$ grid placement. The i-th MR, denoted by MR_i $(1 \le i \le C)$, is connected with N_i^a APs, each of which manages a WLAN [4]. The j-th AP connected with the i-th MR manages a WLAN containing $N_{i,j}^s$ stations, where $1 \le i \le C$ and $1 \le j \le \sum_{i=1}^{C} N_i^a$. There is one queue in the MR and the station. The AP is equipped with two queues at the network layer, one for its uplink-access (i.e., uplink queue) and the other for the downlink-access (i.e., downlink queue) [40].

6.1.2 Network topology

The network topology has a critical impact on the system performance. A number of studies on WMNs have considered an arbitrary topology. For instance, Cao *et al.* [11] investigated the performance of IEEE 802.16 mesh mode scheduler on the arbitrary network topology. Zhou and Mitchell [128] presented an analytical framework for the performance evaluation of CSMA/CA based WMNs with the arbitrary topology. However, Robinson and Knightly [101] showed that an arbitrary network topology is unsuitable for the deployment of WMNs due to doubled node density requirements, while a moderate level of perturbations from ideal grid placement is suitable for the performance study of WMNs. Bisnik and Abouzeid [8] developed an analytical model for WMNs by considering that MRs are distributed in a torus topology, as shown in Fig. 3.1, which is not suitable in practice due to the requirements of wrap-around connections [17]. As a result, the grid placement has been widely adopted as the underlying network topology for the performance study of WMNs [75, 99, 118, 119].

6.1.3 Routing strategies

A random access routing scheme is adopted in this paper [8]. Specifically, packets generated by the station are transferred to the uplink queue of its associated AP in the WLAN, and then are forwarded to the MR. MRs relay the packet over the mesh networking until it reaches the MR where its associated WLAN contains the packet destination. The packets forwarding from the MR to its associated WLAN need to access the downlink queue of AP. The probability that a packet departing from an MR is destined to one of its associated WLANs is σ. The probability that this packet

is forwarded to a neighbouring MR is $(1-\sigma)$. This routing scheme can capture the phenomenon of communication locality exhibited by applications in realistic working environments. The parameter, σ, characterises the degree of communication locality, i.e., the traffic is highly localised for large σ, while a small σ implies unlocalised/random traffic.

6.1.4 Medium access control schemes

To make our model more versatile, we suppose a general MAC scheme [71] with the successful channel-access probability for stations, uplink-access and downlink-access of APs, and MRs, given by $P^s_{i,j,k}$, $P^{au}_{i,j}$, $P^{ad}_{i,j}$, and P^r_i, respectively, where the subscripts, i, j, and k, satisfy the following conditions

$$1 \le i \le C \tag{6.1}$$

$$1 \le j \le \sum_{i=1}^{C} N^a_i \tag{6.2}$$

$$1 \le k \le \sum_{i=1}^{C} \sum_{j=1}^{N^a_i} N^s_{i,j} \tag{6.3}$$

where the subscript i represents the *i*-th MR or the APs and stations associated with the *i*-th MR, the subscript j denotes the *j*-th AP or the stations in the WLAN managed by the *j*-th AP, and the subscript k indicates the *k*-th station. To obtain the probabilities, $P^s_{i,j,k}$, $P^{au}_{i,j}$, $P^{ad}_{i,j}$, and P^r_i, many factors should be considered, including interference, the configuration of physical layer, the number of available channels in the network, the number of available radios of each node, and the operation of MAC protocols [7, 51, 71].

6.2 The analytical model

In this section, an analytical model for the integrated WLANs and multi-hop mesh networks in the presence of bursty traffic with communication locality will be derived. The end-to-end delay experienced by a packet in the whole system is defined as the time between when the first bit of this packet is generated by the station and when the packet is entirely received by its destination. According to the traffic flow in the proposed system as illustrated in Chapter 6.1.3, the delay at stations, uplink-access of APs, MRs, and the downlink-access of APs will first be determined. The expression of end-to-end delay experienced by a packet in the whole system is then derived. A brief summary of key notations used in the derivation of the model is listed in Table 6.1.

Table 6.1: key notations used in the derivation of the model in Chapter 6

C	the number of MRs
N_i^a	the number of APs connected with the *i*-th MR
$N_{i,j}^s$	the *j*-th AP connected with the *i*-th MR manages a WLAN containing $N_{i,j}^s$ stations
$T_{i,j,k}^s$	the inter-arrival time of transmission opportunities for the station
P_v	the probability that there are v packets in the MMPP/M/1 system
$P_{i,j}$	the probability that a packet is forwarded from the queue of MR_i to the queue of MR_j
$\| N_{r_i} \|$, N_{r_i}	the set of neighbouring MRs of MR_i and the number of elements in the set

ε_i	the ratio that a packet is forwarded by MR_i
$\overline{d}$	the average number of hops traversed by a packet in the mesh network
$\overline{D}_{e2e}$	the end-to-end delay experienced by a packet in heterogeneous wireless networks
t_ℓ, t_m	each packet is transmitted within one time slot of length t_ℓ in local area networks and t_m in mesh networks
σ	the probability that a packet departing from an MR is destined to one of its associated WLANs
$\mathbf{Q}_s$, $\mathbf{\Lambda}_s$	parameter matrices of MMPP_s to model the traffic generated by the source node
$\mathbf{Q}_s^o$, $\mathbf{\Lambda}_s^o$	parameter matrices of MMPP_s^o to model the output process from the queue of a station
$\mathbf{Q}_{au}$, $\mathbf{\Lambda}_{au}$	parameter matrices of MMPP_{au} to model the traffic entering the uplink queue of AP
$\mathbf{Q}_{tr}$, $\mathbf{\Lambda}_{tr}$	parameter matrices of MMPP_{tr} to model the total traffic arriving at the WMN
$\mathbf{Q}_r$, $\mathbf{\Lambda}_r$	parameter matrices of MMPP_r to model the traffic arriving at the queue of MR_i
$\mathbf{Q}_{ad}$, $\mathbf{\Lambda}_{ad}$	parameter matrices of MMPP_{ad} to model the traffic entering the downlink-access of AP
MMPP_r^o	modelling the output process from MRs
$P_{i,j,k}^s$, $P_{i,j}^{au}$, $P_{i,j}^{ad}$, P_i^r	the successful channel-access probability for stations, uplink-access and downlink-access of APs, and MRs
$\mu_{i,j,k}^s$, $\mu_{i,j}^{au}$, $\mu_{i,j}^{ad}$, μ_i^r	packet service rate at stations, uplink-access and downlink-access of APs, and MRs
$D_{i,j,k}^s$, $D_{i,j}^u$, $D_{i,j}^{ad}$, D_i^r	the packet delay at the stations, uplink-access and downlink-access of APs, and MRs
$\overline{D}^s$, $\overline{D}^{au}$, $\overline{D}^{ad}$, $\overline{D}^r$	the mean packet delay at the stations, uplink-access and downlink-access of APs, and MRs

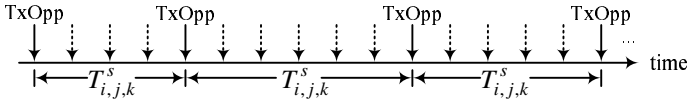

Figure 6.2: Intervals between transmission opportunities (TxOpps) for stations

6.2.1 Packet delay at stations

The packets generated by the stations follow MMPP_s (the subscript s denoting the traffic generated by stations) with the infinitesimal generator $\mathbf{Q}_s$ and rate matrix $\mathbf{\Lambda}_s$. $\mathbf{Q}_s$ and $\mathbf{\Lambda}_s$ can be given by

$$\mathbf{Q}_s = \begin{bmatrix} -\varphi_{s1} & \varphi_{s1} \\ \varphi_{s2} & -\varphi_{s2} \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda}_s = \begin{bmatrix} \lambda_{s1} & 0 \\ 0 & \lambda_{s2} \end{bmatrix} \tag{6.4}$$

To determine the packet delay at the stations, we resort to the MMPP/M/1 queueing system with the packet arrival process modelled by MMPP_s and the packet service rate of the queue which can be determined as follows [71]. Given that the probability that a station is granted a transmission opportunity within one time slot is $P^s_{i,j,k}$ and each packet is transmitted within one time slot of length t_ℓ in the local area networks. Let $T^s_{i,j,k}$ represent the inter-arrival time of transmission opportunities for this station, as depicted in Fig. 6.2, where the solid line with arrow indicates the transmission opportunity for a packet and the dashed line with arrow denotes the failure of packet transmission. The probability that the packet in each station is not transmitted within h time slots can be given by

$$P(T^s_{i,j,k} > ht_\ell) = (1 - P^s_{i,j,k})^h \tag{6.5}$$

Replacing ht_ℓ by t, we have

$$P(T_{i,j,k}^{s} > t) = (1 - P_{i,j,k}^{s})^{t/t_\ell} \tag{6.6}$$

Since binomial distribution can be approximated by Poisson distribution with appropriate parameters [38], the arrival process of transmission opportunities is approximated as a Poisson process. In other words, an exponential distribution with mean $1/\mu_{i,j,k}^{s}$ is used to model the inter-arrival time of transmission opportunities, $T_{i,j,k}^{s}$. Therefore, we have

$$\mu_{i,j,k}^{s} \approx \frac{\ln(1 - P_{i,j,k}^{s})^{-1}}{t_\ell} \tag{6.7}$$

The packet delay, $D_{i,j,k}^{s}$, at the station can be obtained by modelling the station as an MMPP/M/1 queueing system with the packet arrival process modelled by MMPP_s and the packet service rate, $\mu_{i,j,k}^{s}$. Therefore, $D_{i,j,k}^{s}$ can be determined by the packet waiting time in the system as follows:

$$D_{i,j,k}^{s} = \frac{L_{i,j,k}^{s}}{\overline{\lambda}_s} \tag{6.8}$$

where $L_{i,j,k}^{s}$ denotes the mean system length of MMPP/M/1 queueing system and $\overline{\lambda}_s$ is the mean arrival rate of MMPP_s given by Eq. (2.11). In order to compute $L_{i,j,k}^{s}$, we need to derive P_v $(0 \le v \le \infty)$, which denotes the probability that there are v packets in the MMPP/M/1 queueing system. Based on Eqs. (3.46) and (3.47), P_v can be given by

$$P_v = \boldsymbol{\pi}(\mathbf{I} - \mathbf{R})\mathbf{R}^{v}\mathbf{e}, \quad 0 \le v \le \infty \tag{6.9}$$

where $\boldsymbol{\pi}$ is the steady state vector of MMPP_s and $\mathbf{e}$ is the column unit vector of

length 2. $\mathbf{I}$ is the identity matrix of size 2. The factor, $\mathbf{R}$, can be obtained by solving the quadratic matrix equation

$$\mathbf{A} + \mathbf{RB} + \mathbf{R}^2\mathbf{C} = 0 \tag{6.10}$$

with $\mathbf{A} = \mathbf{\Lambda}_s$, $\mathbf{B} = \mathbf{Q}_s - \mathbf{\Lambda}_s - \mathbf{I}/\mu^s_{i,j,k}$, and $\mathbf{C} = \mathbf{I}/\mu^s_{i,j,k}$. The mean system length, $L^s_{i,j,k}$, can be obtained as

$$L^s_{i,j,k} = \sum_{v=0}^{\infty} vP_v = \boldsymbol{\pi}\mathbf{R}(\mathbf{I}-\mathbf{R})^{-1}\mathbf{e} \tag{6.11}$$

The mean packet delay, $\overline{D}^s$, at the station can be determined by

$$\overline{D}^s = \frac{\sum_{i=1}^{C}\sum_{j=1}^{N_i^a}\sum_{k=1}^{N^s_{i,j}} D^s_{i,j,k}}{\sum_{i=1}^{C}\sum_{j=1}^{N_i^a} N^s_{i,j}} \tag{6.12}$$

6.2.2 Packet delay at the uplink-access of access points

Each packet departing from the stations needs to enter the uplink queue of its associated AP. To obtain the arrival process of the traffic entering the uplink queue of APs, the output process from the queue of the stations needs to be derived. In [29], the authors proposed an approximation to modelling the output process from an MMPP/M/1 queue. Let MMPP^o_s represent the output process from the queue of a station. The output process is obtained by matching the inter-departure time moments. Let T_υ denote the time between the υ-th and $(\upsilon+1)$-th departures of the output process. From [30], the moments of inter-departure time can be given by

$$E[T_\upsilon^n] = (-1)^n\left[\sum_{\upsilon=0}^{n-1}\frac{n!}{\upsilon!}H^{(\upsilon)}(0)\mathbf{x}_0\mathbf{B}^{-(n-\upsilon)}(0)\mathbf{e}\right] + (-1)^n H^{(n)}(0) \tag{6.13}$$

where $\mathbf{B}(0) = \mathbf{Q}_s - \mathbf{\Lambda}_s$ and $\mathbf{x}_0$ is the probability that the queue is empty observed at the departure instants and can be found in [31]. $H(s)$ is the Laplace-Stieltjes transform of the service time distribution and can be given by

$$H(s) = \frac{\frac{1}{\mu_{i,j}^{au}}}{\frac{1}{\mu_{i,j}^{au}} + s} \tag{6.14}$$

According to the moment matching scheme [29], the infinitesimal generator, $\mathbf{Q}_s^o$, and rate matrix, $\mathbf{\Lambda}_s^o$, of MMPP_s^o can be obtained.

Let MMPP_{au} denote the traffic entering the uplink queue of the AP. Based on the flow conservation law, i.e., the total number of packets departing from all stations in a WLAN must be equal to the total number of packets entering the uplink queue of their associated AP, MMPP_{au} is the superposition of $N_{i,j}^s$ traffic flows modelled by MMPP_s^o. Based on Eqs. (3.17)-(3.34), the infinitesimal generator, $\mathbf{Q}_{au}$, and rate matrix, $\mathbf{\Lambda}_{au}$, of MMPP_{au} can be readily obtained.

The packet delay at the uplink-access of AP, $D_{i,j}^u$, can be determined by modelling the uplink-access as an MMPP/M/1 queueing system with the arrival process, MMPP_{au}, and the service rate, $\mu_{i,j}^{au}$. $D_{i,j}^u$ can be obtained based on Eqs. (6.8)-(6.11). According to the determination of MMPP_s^o, the output process from the uplink queue of APs, MMPP_{au}^o, can be obtained. The mean delay, $\overline{D}^{au}$, experienced by a packet at the uplink-access of AP can be obtained by virtue of Eq. (6.12).

6.2.3 Packet delay at mesh routers

Let MMPP_{tr} denote the total traffic arriving at the WMN. MMPP_{tr} can be obtained by the superposition of $\sum_{i=1}^{C} N_i^a$ traffic flows modelled by MMPP_{au}^o. According to the derivation of MMPP_{au}, the infinitesimal generator, $\mathbf{Q}_{tr}$, and rate matrix, $\mathbf{\Lambda}_{tr}$, of MMPP_{tr} can be determined.

Since each packet leaving the queue of an MR has the probability, $(1-\sigma)$, to be randomly forwarded to one of its neighbouring MRs, the probability that a packet is forwarded from the queue of MR_i $(1 \le i \le C)$ to the queue of MR_j, $P_{i,j}$, can be computed as follows

$$P_{i,j} = \begin{cases} \dfrac{1-\sigma}{N_{r_i}} & j \in \| N_{r_i} \| \\ 0 & \text{otherwise} \end{cases} \tag{6.15}$$

where $\| N_{r_i} \|$ is the set of neighbouring MRs of MR_i and N_{r_i} is the number of elements in the set. These two quantities can be obtained according to Fig. 6.1.

The ratio, ε_i, that a packet is forwarded by MR_i is given by the sum of the probability that a packet enters the queue of MR_i from its associated WLANs and the probability that a packet is routed to the queue of MR_i after completing the service from its neighbouring MRs. Thus, ε_i can be given by

$$\varepsilon_i = \frac{\sum_{j=1}^{N_i^a} N_{i,j}^s}{\sum_{i=1}^{C}\sum_{j=1}^{N_i^a} N_{i,j}^s} + \sum_{j \in \| N_{r_i} \|} P_{j,i}\varepsilon_j \tag{6.16}$$

Let $\boldsymbol{\varphi}$ and $\mathbf{A}$ denote row vectors, where

$$\boldsymbol{\varphi} = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_C] \quad \text{and}$$

$$\mathbf{A} = \left[\frac{\sum_{j=1}^{N_1^a} N_{1,j}^s}{\sum_{i=1}^{C}\sum_{j=1}^{N_i^a} N_{i,j}^s}, \frac{\sum_{j=1}^{N_2^a} N_{2,j}^s}{\sum_{i=1}^{C}\sum_{j=1}^{N_i^a} N_{i,j}^s}, ..., \frac{\sum_{j=1}^{N_C^a} N_{C,j}^s}{\sum_{i=1}^{C}\sum_{j=1}^{N_i^a} N_{i,j}^s} \right]$$

Let $\mathbf{P}$ represent the forwarding probability of $P_{i,j}$. Therefore, Eq. (6.16) can be re-written in the form

$$\boldsymbol{\varphi} = \mathbf{A}(\mathbf{I} - \mathbf{P})^{-1} \tag{6.17}$$

where $\mathbf{I}$ is the identity matrix of order C. Since $(\mathbf{I} - \mathbf{P})$ is non-singular, Eq. (6.17) has a unique solution. Let MMPP_r represent the traffic arriving at the queue of MR_i. Based on Eqs. (3.17)-(3.34), the infinitesimal generator, $\mathbf{Q}_r$, and rate matrix, $\mathbf{\Lambda}_r$, of MMPP_r can be readily determined. The service rate of the queue in MR_i, μ_i^r, can be obtained according to Eqs. (6.5)-(6.7).

The delay experienced by a packet to cross MR_i, D_i^r, can be obtained by modelling MR as an MMPP/M/1 queueing system with the arrival process, MMPP_r, and the service rate, μ_i^r, and can be determined by virtue of Eqs. (6.8)-(6.11). According to the derivation of MMPP_s^o, the output process from the MR, MMPP_r^o, can be obtained. Based on Eq. (6.12), the mean packet delay, $\overline{D}^r$, at an MR can be determined.

6.2.4 Packet delay at the downlink-access of access points

Recall that packets departing from MR_i have the probability, σ, to randomly choose one of its associated APs as its next hop. The traffic entering the downlink-access of AP, MMPP_{ad}, can be determined by the splitting of MMPP_r^o with the

splitting probability, σ. Based on Eq. (3.18), the infinitesimal generator, $\mathbf{Q}_{ad}$, and rate matrix, $\mathbf{\Lambda}_{ad}$, of MMPP_{ad} can be easily obtained.

The service rate of downlink queue in an AP, $\mu_{i,j}^{ad}$, can be determined according to Eqs. (6.5)-(6.7), and the delay experienced by a packet to cross the downlink-access of AP, $D_{i,j}^{ad}$, can be obtained by virtue of Eqs. (6.8)-(6.11). The mean packet delay, $\overline{D}^{ad}$, at the downlink-access of AP can be obtained according to Eq. (6.12).

6.2.5 End-to-end delay in the whole system

The average number of hops traversed by a packet in the mesh network, $\bar{d}$, can be computed as follows

$$\bar{d} = \sum_{v=1}^{\infty} v(1-\sigma)^{v-1}\sigma = \frac{1}{\sigma} \tag{6.19}$$

The end-to-end delay, $\overline{D}_{e2e}$, is equal to the sum of the packet transmission time, the queueing delays for all the intermediate APs and MRs, and the delay at the source station. Therefore, $\overline{D}_{e2e}$ can be expressed as

$$\overline{D}_{e2e} = \bar{d}t_m + 4t_\ell + \overline{D}^{au} + \overline{d}\overline{D}^{r} + \overline{D}^{ad} + \overline{D}^{s} \tag{6.20}$$

6.3 Validation of the model

The accuracy of the analytical model is validated by means of a discrete event-driven simulator based on OMNeT++ [89]. Each simulation experiment was run until the network reached its steady state and collected based on Eqs. (3.51) and (3.52). The

Table 6.2: System configuration parameters for integrated WLANs and WMN

Simulation Cases	N_i^a (Number of APs)	$N_{i,j}^s$ (Number of Stations)
Case 1	$N_i^a = 3,\ i \in [1, C/2]$	$N_{i,j}^s = 5,\ i \in [1, C/2],\ j \in [1, 3]$
	$N_i^a = 3,\ i \in [C/2+1, C]$	$N_{i,j}^s = 5,\ i \in [C/2+1, C],\ j \in [1, 3]$
Case 2	$N_i^a = 3,\ i \in [1, C/2]$	$N_{i,j}^s = 4,\ i \in [1, C/2],\ j \in [1, 3]$
	$N_i^a = 2,\ i \in [C/2+1, C]$	$N_{i,j}^s = 6,\ i \in [C/2+1, C],\ j \in [1, 2]$

arrivals of messages generated by the station are modelled by MMPP_s with the infinitesimal generator $\mathbf{Q}_s$ and rate matrix $\mathbf{\Lambda}_s$. The message destinations are non-uniformly distributed over the network nodes according to communication locality.

Extensive simulation experiments have been performed to validate the accuracy of the analytical model. For the sake of specific illustration, the following cases are considered with the system parameters shown in Table 6.2, where N_i^a is the number of APs associated with the *i*-th MR and $N_{i,j}^s$ is the number of stations in the *j*-th WLAN which is connected with the *i*-th MR. Case 1 and Case 2 consider homogeneous and heterogeneous number of APs connected with one MR and number of stations in a WLAN, respectively. Without loss of generality, we consider a system with 81 MRs uniformly distributed in a 9×9 grid placement. The system parameters are set as follows:

- the channel capacities in the mesh network and local area networks are 54 Mb/s and 11 Mb/s, respectively;
- the packet size is set to 8000 bits;
- the length of a time slot is set to be the amount of airtime needed to transmit one

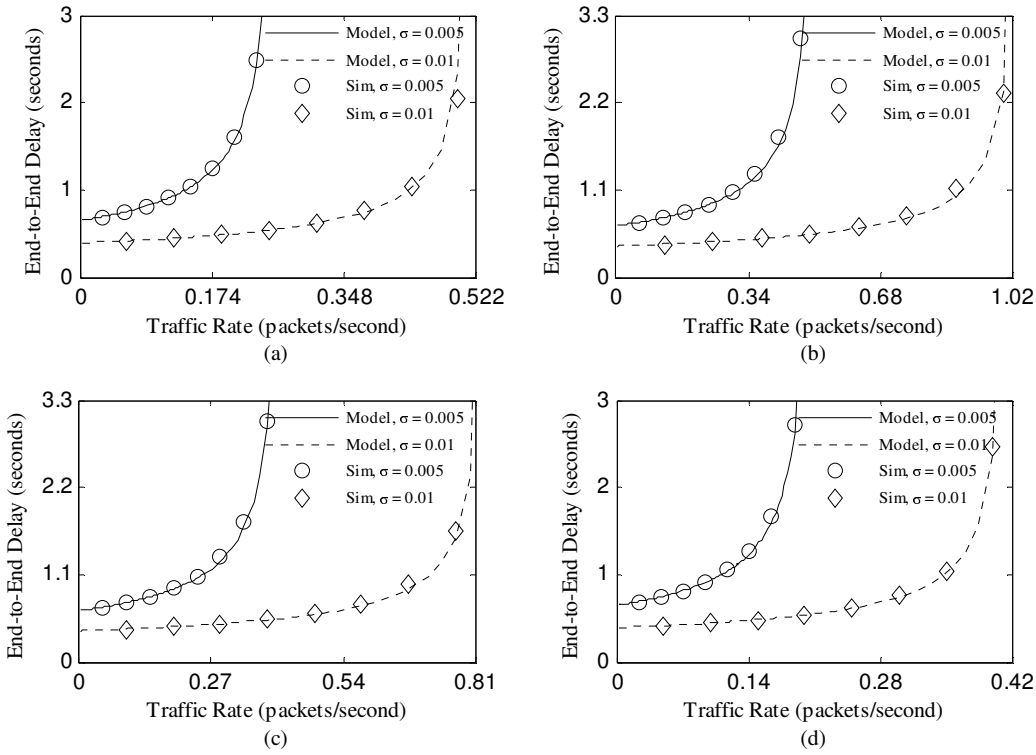

Figure 6.3: End-to-end delay predicted by the model and simulation with Case 1 in Table 6.2: (a) φ_{s1} = 0.09, φ_{s2} = 0.06, (b) φ_{s1} = 0.08, φ_{s2} = 0.02, (c) φ_{s1} = 0.09, φ_{s2} = 0.03 and (d) φ_{s1} = 0.004, φ_{s2} = 0.004

packet, i.e., t_m = 1.48e-4 s in the mesh network and t_ℓ = 7.27e-4 s in local area networks;

- the successful channel-access probabilities for stations, uplink-access and downlink-access of APs, and MRs are set as $P^s_{i,j,k}$ = 9e-3, $P^{au}_{i,j}$ = 2e-2, $P^{ad}_{i,j}$ = 4e-2, and P^r_i = 6e-2, respectively;
- the communication locality is set to σ = 0.1 and 0.15, representing different degrees of communication locality;
- the infinitesimal generator, $\mathbf{Q}_s$, of MMPP$_s$ are set as follows, representing different degrees of traffic burstiness and correlations:

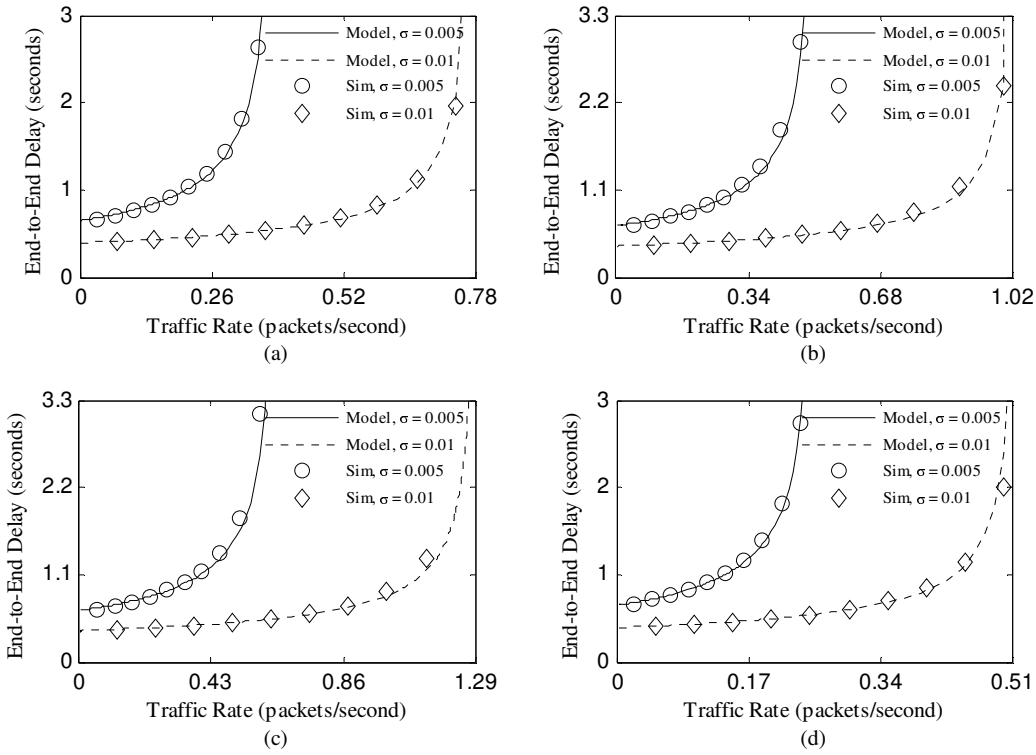

Figure 6.4: End-to-end delay predicted by the model and simulation with Case 2 in Table 6.2: (a) φ_{s1} = 0.07, φ_{s2} = 0.035, (b) φ_{s1} = 0.06, φ_{s2} = 0.02, (c) φ_{s1} = 0.04, φ_{s2} = 0.01 and (d) φ_{s1} = 0.09, φ_{s2} = 0.09

$$\begin{bmatrix} -0.09 & 0.09 \\ 0.06 & -0.06 \end{bmatrix}, \begin{bmatrix} -0.08 & 0.08 \\ 0.02 & -0.02 \end{bmatrix}, \begin{bmatrix} -0.09 & 0.09 \\ 0.03 & -0.03 \end{bmatrix}, \begin{bmatrix} -0.004 & 0.004 \\ 0.004 & -0.004 \end{bmatrix}$$

$$\begin{bmatrix} -0.07 & 0.07 \\ 0.035 & -0.035 \end{bmatrix}, \begin{bmatrix} -0.06 & 0.06 \\ 0.02 & -0.02 \end{bmatrix}, \begin{bmatrix} -0.04 & 0.04 \\ 0.01 & -0.01 \end{bmatrix}, \text{ and}$$

$$\begin{bmatrix} -0.09 & 0.09 \\ 0.09 & -0.09 \end{bmatrix}.$$

Figs. 6.3 and 6.4 show the analytical results predicted by the model plotted against those provided by the simulator in the integrated wireless networks. In these figures, the horizontal axis represents the traffic rate, λ_{s1}, at which a station injects messages into the network when the MMPP$_s$ is at state 1, while the vertical axis denotes the end-to-end delay obtained from the above model. For the sake of clarity

of the figures, the arrival rate, λ_{s2}, at state 2 is deliberately set to zero; otherwise the three-dimensional graphs will be used to represent the results. These figures reveal that the results obtained from the analytical model closely match those obtained from the simulation. The tractability and accuracy of the analytical model make it a practical and cost-effective tool to evaluate and optimise the performance of the integrated wireless networks in the presence of bursty traffic with communication locality.

6.4 Performance analysis

In this section, the maximum achievable throughput in the integrated wireless networks will first be derived and the impact of bursty traffic with communication locality on network performance will then be investigated.

6.4.1 Maximum achievable throughput

The maximum achievable throughput, which can reflect the capacity of a wireless network, is defined as the maximum value of the arrival rate of packets at stations where the end-to-end delay, $\overline{D}_{e2e}$, remains finite. For the end-to-end delay to be finite, the delay experienced by a packet at the stations, uplink-access and down-access of APs, and the MRs should be finite. To this end, the utilisation factor of the queue in the stations, uplink queue in the APs, the queue in the MR_i $(1 \le i \le C)$, and the downlink queue in the AP_i must be less than one, i.e.,

$$\frac{\overline{\lambda}_x}{\mu_x} < 1, \quad x \in \{s, au, r, ad\} \tag{6.21}$$

For $x = s$, we have

$$\overline{\lambda}_s < \mu^s_{i,j,k} = t_\ell^{-1} \ln(1 - P^s_{i,j,k})^{-1} = T^s_{i,j,k} \tag{6.22}$$

$$T_s^{max} \overset{\Delta}{=} \min\left\{ T^s_{i,j,k},\, i \in [1, C],\; j \in \left[1, \sum_{i=1}^{C} N_i^a\right],\, k \in \left[1, \sum_{i=1}^{C} \sum_{j=1}^{N_i^a} N^s_{i,j}\right] \right\} \tag{6.23}$$

For $x = au$, we have

$$\overline{\lambda}_s < \frac{\mu^{au}_{i,j}}{N^s_{i,j}} = \frac{t_\ell^{-1} \ln(1 - P^{au}_{i,j})^{-1}}{N^s_{i,j}} = T^{au}_{i,j} \tag{6.24}$$

$$T_{au}^{max} \overset{\Delta}{=} \min\left\{ T^{au}_{i,j},\; i \in [1, C],\; j \in \left[1, \sum_{i=1}^{C} N_i^a\right] \right\} \tag{6.25}$$

According to the derivation of MMPP_r, characterising the traffic arriving at the queue of MR_i, we can obtain $\overline{\lambda}_r = \varepsilon_i^r \overline{\lambda}_s$, where ε_i^r is a constant coefficient. For $x = r$, we have

$$\overline{\lambda}_s < \frac{\mu_i^r}{\varepsilon_i^r} = \frac{t_m^{-1} \ln(1 - P_i^r)^{-1}}{\varepsilon_i^r} = T_i^r \tag{6.26}$$

$$T_r^{max} \overset{\Delta}{=} \min\{T_i^r,\; i \in [1, C]\} \tag{6.27}$$

For $x = ad$, we can write

$$\overline{\lambda}_s < \frac{\mu^{ad}_{i,j}}{\varepsilon^{ad}_{i,j}\sigma} = \frac{t_\ell^{-1} \ln(1 - P^{ad}_{i,j})^{-1}}{\varepsilon^{ad}_{i,j}\sigma} = T^{ad}_{i,j} \tag{6.28}$$

where, similar to ε_i^r used in Eq. (6.26), ε_i^{ad} is a constant coefficient to satisfy $\overline{\lambda}_{ad} = \varepsilon^{ad}_{i,j} \overline{\lambda}_s$.

$$T_{ad}^{max} \overset{\Delta}{=} \min\left\{ T^{ad}_{i,j},\; i \in [1, C],\; j \in \left[1, \sum_{i=1}^{C} N_i^a\right] \right\} \tag{6.29}$$

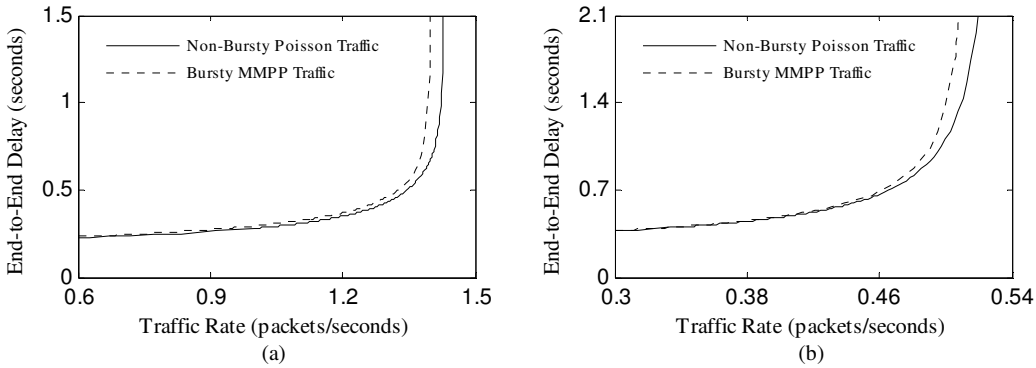

Figure 6.5: End-to-end delay predicted by the derived analytical model and the model under the Poisson assumption with (a) $C = 10$, $\sigma = 0.05$, $\varphi_{s1} = 0.08$, $\varphi_{s2} = 0.06$ and (b) $C = 6$, $\sigma = 0.02$, $\varphi_{s1} = 0.007$, $\varphi_{s2} = 0.007$

Therefore, the maximum achievable throughput, T^{max}, can be expressed as

$$T^{max} = \min(T_s^{max}, T_{au}^{max}, T_r^{max}, T_{ad}^{max}) \tag{6.30}$$

6.4.2 The impact of bursty traffic on network performance

Having validated the accuracy of the analytical model, we will use it as a cost-effective tool to investigate the impact of bursty traffic on the performance of integrated wireless networks. To this end, Fig. 6.5 depicts the results predicted by the model under the bursty MMPP against those obtained from the model under non-bursty Poisson process (i.e., the assumption made in our previous work [83]) as a function of the mean arrival rate in 10×10 and 6×6 grid placement with $N_i^a = 3$ for $1 \le i \le C/2$, $N_{i,j}^s = 4$ for $1 \le i \le C/2$ and $1 \le j \le 3$; and $N_i^a = 2$ for $C/2+1 \le i \le C$, $N_{i,j}^s = 6$ for $C/2+1 \le i \le C$ and $1 \le j \le 2$. The channel capacity in the WMN and individual wireless networks is 54 Mb/s and 11 Mb/s, respectively, the packet size is set to 8000 bits, and the successful channel-access probabilities for the stations, the

uplink-access of APs, the MRs, the downlink-access of APs are set to be $P_{i,j,k}^{s}$ = 9e-3, $P_{i,j}^{au}$ = 2e-2, $P_{i,j}^{ad}$ = 4e-2, and P_{i}^{r} = 6e-2, $(1 \le i \le C)$, respectively. It can be seen from the figure that the model based on the Poisson traffic under-estimates the end-to-end delay when the network is subject to the bursty traffic. In other words, the bursty traffic can degrade the system performance seriously since the end-to-end delay experienced by a packet in integrated wireless networks increases and the maximum achievable throughput decreases. The result is consistent with that is concluded in Chapter 3.5.1 and Chapter 4.5.2. Our finding emphasises the importance of using realistic models to study and optimise the performance of integrated wireless networks.

6.4.3 The impact of communication locality on network performance

Due to the important phenomenon of communication locality in wireless networks, in what follows, we will investigate its impact on the performance (i.e., end-to-end delay and maximum achievable throughput) of the integrated WLANs and multi-hop mesh networks. To this end, Fig. 6.6 depicts the results obtained from the analytical model with varying degrees of communication locality in the 6×6 grid placement with $N_{i}^{a} = 3$ for $1 \le i \le C/2$, $N_{i,j}^{s} = 4$ for $1 \le i \le C/2$ and $1 \le j \le 3$; and $N_{i}^{a} = 2$ for $C/2+1 \le i \le C$, $N_{i,j}^{s} = 6$ for $C/2+1 \le i \le C$ and $1 \le j \le 2$. The channel capacity in the WMN and individual wireless networks is 54 Mb/s and 11 Mb/s, respectively, the packet size is set to 8000 bits, and the successful channel-access probabilities for the stations, the uplink-access of APs, the MRs, the downlink-access of APs are set to be

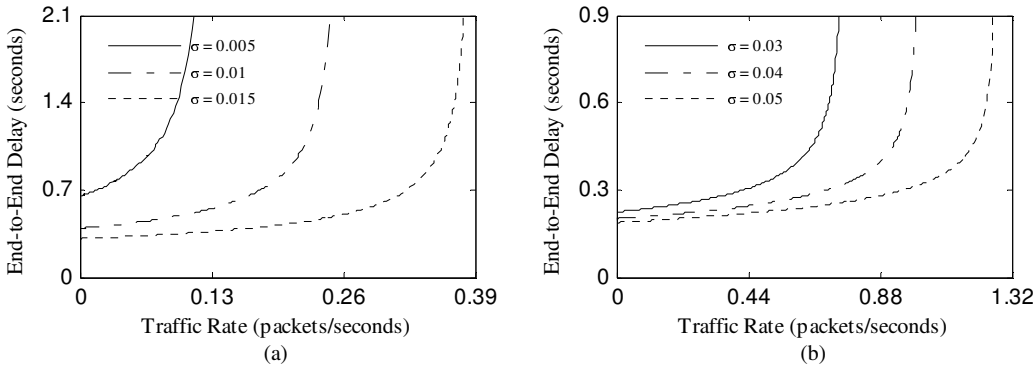

Figure 6.6: End-to-end delay predicted by the analytical model with φ_{s1} = 0.08, φ_{s2} = 0.04: (a) σ = 0.005, 0.01 and 0.015 and (b) σ = 0.03, 0.04 and 0.05

$P_{i,j,k}^{s}$ = 9e-3, $P_{i,j}^{au}$ = 2e-2, $P_{i,j}^{ad}$ = 4e-2, and P_{i}^{r} = 6e-2, $(1 \le i \le C)$, respectively. It can be seen from the figure that increasing the communication locality improves the network performance since the end-to-end delay experienced by a packet decreases. This is because the average number of hops traversed by a packet drops when the locality rises, as shown by Eq. (6.19). Moreover, we can find that the maximum achievable throughput increases as the communication locality goes up, i.e., a higher degree of communication locality can enhance the network capacity, which is very important in wireless networks. Examining Fig. 6.6(b) reveals the same results. The result is consistent with that is concluded in Chapter 4.5.1. The above analysis stresses the great need for the development of new wireless networking technologies that can exploit the communication locality.

6.5 Conclusions

This chapter has developed a new analytical model to investigate the end-to-end delay in a WMN interconnecting multiple WLANs based on a random access routing strategy and a general MAC scheme. The model can capture the bursty nature of

packet arrival process and the important phenomenon of communication locality exhibited in wireless networks. Extensive simulation experiments have been conducted to validate the accuracy of the analytical model. The results have revealed that the end-to-end delay predicted by the model closely match those obtained from the simulation. The analytical model has been used to obtain the maximum achievable throughput which reflects the capacity of a wireless network. Moreover, the model has been used as a cost-effective tool to investigate the impact of bursty traffic with communication locality on network performance. The analytical results have shown that the bursty traffic can degrade the system performance seriously and the communication locality can reconcile the degrading effects caused by the bursty traffic and improves the performance of integrated wireless networks.

Chapter 7

Analytical Modelling of Integrated WLANs and Internet-access Mesh Networks under Bursty Traffic

The prediction of the end-to-end performance measures for peer-to-peer communications (i.e., stations communicate with each other) in the integrated WLANs and WMN have been investigated in Chapter 6. However, the delay and throughput analysis in the integrated WLANs and Internet-access mesh networks (i.e., stations communicate with the gateway) has seldom been studied. To fill this gap, this chapter presents a new analytical model to investigate the end-to-end delay and throughput in integrated WLANs and Internet-access mesh networks in the presence of bursty traffic. To this end, the network entities (i.e., stations, APs, and MRs) in the hybrid topology are modelled as finite capacity queueing systems and derive the expressions of the queue length, throughput, and packet waiting time by virtue of queueing theory. Extensive simulation experiments are carried out to validate the accuracy of the analytical model. The analytical model is then adopted as a cost-effective tool to investigate the bottleneck problem in the integrated wireless networks.

The rest of this chapter is organised as follows. Section 7.1 presents the

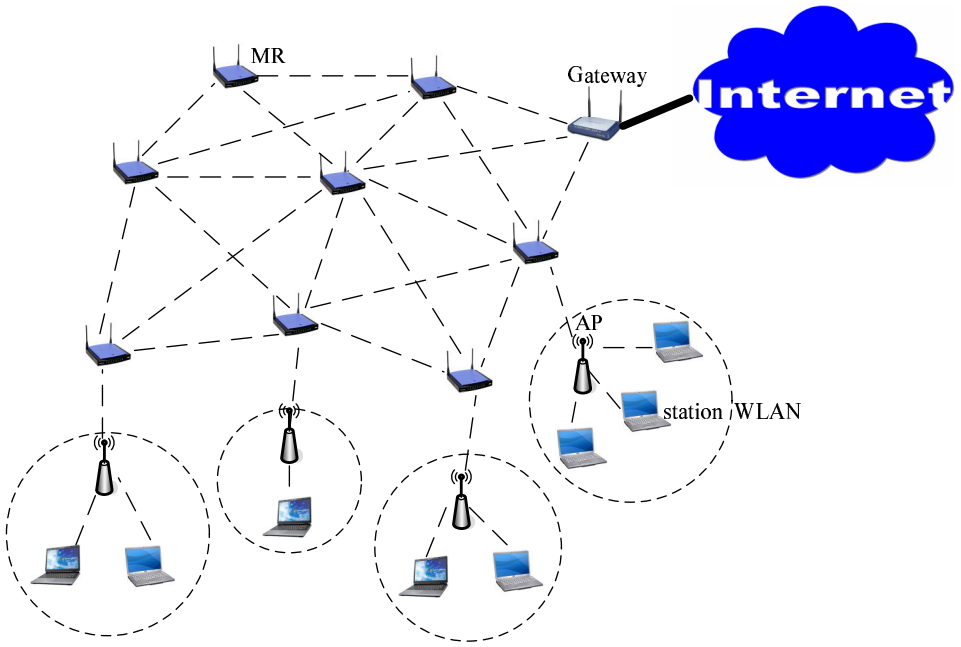

Figure 7.1: Architecture of the integrated WLANs and Internet-access mesh networks

architecture of integrated WLANs and Internet-access mesh networks. Section 7.2 derives the analytical model to investigate the end-to-end delay and throughput in the integrated wireless networks. The accuracy of the model is validated in Section 7.3. Section 7.4 carries out performance analysis. Finally, Section 7.5 concludes this study.

7.1 The architecture of integrated WLANs and Internet-access mesh networks

This chapter focuses on the integrated network architecture, as depicted in Fig. 7.1, which consists of *C* WLANs and a WMN acting as the multi-hop backhaul network to provide the Internet access. The dashed lines indicate the possible communications between the pairs of entities. Each MR is connected to one AP that manages a specific WLAN [4, 25, 57]. The backhaul network has one gateway connected to the

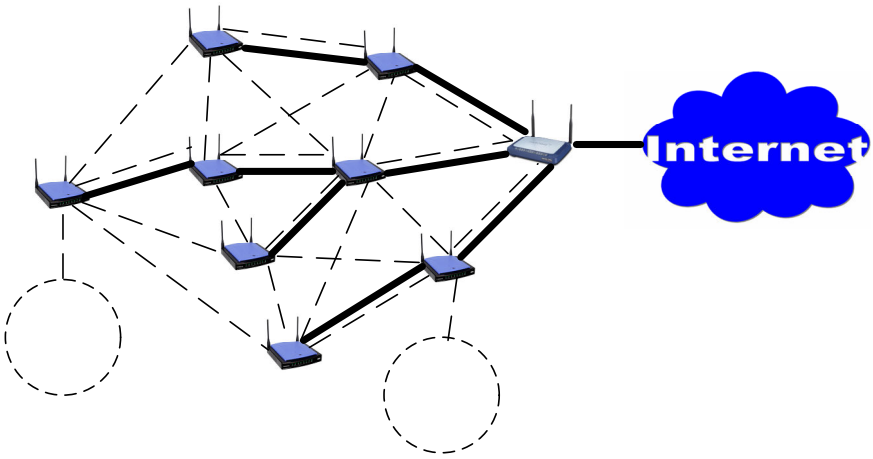

Figure 7.2: The tree-based logical structure of the traffic pattern in integrated WLANs and Internet-access mesh networks

Internet. Each WLAN contains N stations; each station, AP, and MR is equipped with one finite-capacity buffer.

Packets are transmitted in a multi-hop manner towards or from the gateway in the WMN. For clarity, we focus on the upstream traffic, i.e., packets generated by stations are transferred to the associated AP and then are forwarded to the gateway via MRs in a multi-hop fashion. Due to the highly connected feature of WMNs [4], the shortest path routing scheme is used to relay packets in the integrated wireless networks. The traffic streams in the network form a tree-based logical structure as shown in Fig. 7.2. The bold lines represent the routing path in the network. The circles with dashed arced lines denote a set of WLANs.

7.2 The analytical model

In what follows, the delay and throughput in the integrated WLANs and Internet-access mesh networks will be analysed. The traffic generated by the stations follows MMPP_s (the subscript s denoting the traffic generated by stations) with the

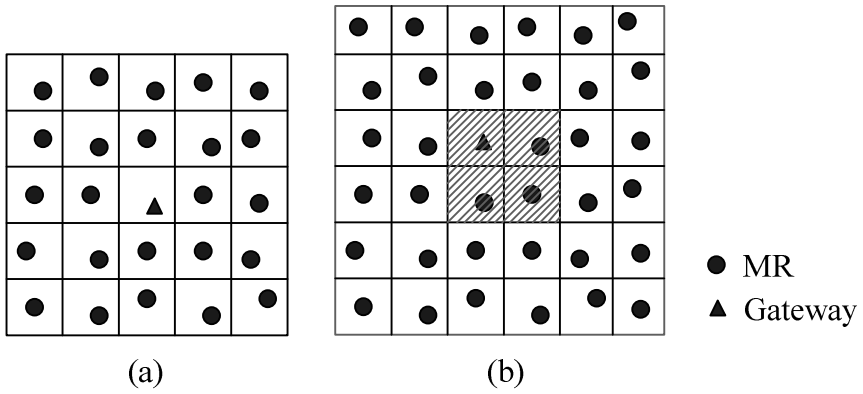

Figure 7.3: MRs in the grid topology: (a) 5×5 grid and (b) 6×6 grid

infinitesimal generator $\mathbf{Q}_s$ and rate matrix $\mathbf{\Lambda}_s$. $\mathbf{Q}_s$ and $\mathbf{\Lambda}_s$ can be given by

$$\mathbf{Q}_s = \begin{bmatrix} -\varphi_{s1} & \varphi_{s1} \\ \varphi_{s2} & -\varphi_{s2} \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda}_s = \begin{bmatrix} \lambda_{s1} & 0 \\ 0 & \lambda_{s2} \end{bmatrix} \tag{7.1}$$

The packet is transmitted within one time slot of length t_ℓ in the WLAN and length t_m in the multi-hop WMN. As a moderate level of perturbations from the ideal grid placement of MRs is suitable for the performance study of WMNs [101], we consider an $n \times n$ grid topology where MRs are uniformly distributed, as shown in Fig. 7.3. For the purpose of illustration, the address (x, y), $0 \le x, y \le n-1$, is used to denote the location of MRs in the grid topology, where x and y denote the row and column order of MRs.

As shown in Fig. 7.2, the tree-based logical structure causes the traffic distributed non-uniformly across the network. Let us define the "*i*-hop MRs" as the MRs located at i hops away from the gateway. To maximise the capacity and minimise the average delay, the gateway is deployed at the centre of the network [4, 71]. To address this tree-based logical structure of traffic patterns, in what follows, the number of MRs located at i hops away from the gateway will first be derived, where $1 \le i \le h_{max}$ and h_{max} denotes the maximum possible hop-count distance from

the MR to the gateway. Following the order of the traffic flows to cross the integrated wireless networks, the performance measures (i.e., delay and throughput) at stations, APs, and MRs will then be determined. Finally, the expressions of the end-to-end delay and throughput for the integrated WLANs and Internet-access mesh networks will be obtained.

To make our model more versatile, we suppose a general MAC scheme [71] with the successful channel-access probabilities for stations, APs, and MR_i $(1 \le i \le h_{max})$, given by P_s, P_a, and P_{r_i}, respectively. The specification of this type of MAC scheme can be found in Chapter 6.1.4. A brief summary of key notations used in the derivation of the model is listed in Table 7.1.

Table 7.1: Key notations used in the derivation of the model in Chapter 7

C	the number of MRs
N_i	the number of the i-hop MRs
N	the number of stations in a WLAN
h_{max}	the maximum possible hop-count distance from the MR to the gateway
$P_{v,j}$	the probability that there are v $(0 \le v \le K_s)$ packets in the queue of station and the underlying Markov chain of the MMPP_s is at state j $(j = 1, 2)$
P_v	the probability that there are v packets in queue of the station
t_ℓ, t_m	each packet is transmitted within one time slot of length t_ℓ in local area networks and t_m in mesh networks
ω	packet size
$\mathbf{Q}_s$, $\mathbf{\Lambda}_s$	parameter matrices of MMPP_s to model the traffic generated by the station

$\mathbf{Q}_s^e$, $\mathbf{\Lambda}_s^e$	parameter matrices of MMPP_s^e to model the effective traffic entering the queue of a station
$\mathbf{Q}_s^o$, $\mathbf{\Lambda}_s^o$	parameter matrices of MMPP_s^o to model the output process from the queue of a station
$\mathbf{Q}_a$, $\mathbf{\Lambda}_a$	parameter matrices of MMPP_{au} to model the traffic entering the uplink queue of an AP
$\mathbf{Q}_a^e$, $\mathbf{\Lambda}_a^e$	parameter matrices of MMPP_a^e to model the effective traffic entering the queue of an AP
$\mathbf{Q}_a^o$, $\mathbf{\Lambda}_a^o$	parameter matrices of MMPP_a^o to model the output process from the queue of an AP
MMPP_{r_i}	traffic arriving at the queue of MRs located i hops away from the gateway
$\text{MMPP}_{r_i}^e$	the effective traffic entering the queue of MRs located i hops away from the gateway
$\text{MMPP}_{r_i}^o$	the output process from the queue of MRs located i hops away from the gateway
D_i^{e2e}	the delay experienced by a packet from its source station to the gateway
T_i	the average number of packets generated by each station in a WLAN, which is associated with MR located at i hops away from the gateway, successfully received by the gateway
T	the total number of packets received by the gateway
D_{e2e}	the average end-to-end delay experienced by the packet in the integrated WLANs and Internet-access mesh networks
P_s, P_a, P_{r_i}	the successful channel-access probability for stations, APs, and MR_i $(1 \le i \le h_{max})$
μ_s, μ_a, μ_{r_i}	packet service rate at stations, APs, and MR_i
Pb_s, Pb_a, Pb_{r_i}	packet blocking probability at stations, APs, and MR_i
$\overline{\lambda}_s^e$, $\overline{\lambda}_a^e$, $\overline{\lambda}_{r_i}^e$	the mean arrival rate of MMPP_s^e, MMPP_a^e, and $\text{MMPP}_{r_i}^e$
D_s, D_a, D_{r_i}	the packet delay at stations, APs, and MR_i

7.2.1 Number of the *i*-hop mesh routers

Let N_i denote the number of the *i*-hop MRs. We need distinguish the situations when n is odd or even to calculate N_i in the grid topology. When n is odd, we have

$$N_i = 8 \times i \tag{7.2}$$

When n is even, the gateway can be placed at the four alternative locations as shown in the shadow area in Fig. 7.3 (b). Thus, N_i can be given by

$$N_i = \begin{cases} 8 \times i & i \le \min(x, y) \\ 2 \times n - 1 & \text{otherwise} \end{cases} \tag{7.3}$$

7.2.2 Packet delay and throughput at stations

Recall that the packets arriving at the queue in stations follow MMPP_s with the infinitesimal generator $\mathbf{Q}_s$ and rate matrix $\mathbf{\Lambda}_s$. To derive the packet delay and throughput at stations, we resort to the $\text{MMPP}_s/\text{M}/1/K_s$ queueing theory with finite capacity K_s in the queue of stations. Since the successful channel-access probability for the stations is given by P_s, the service rate, μ_s, of the $\text{MMPP}_s/\text{M}/1/K_s$ queue can be easily obtained based on Eqs. (6.5)-(6.7).

Given the finite queue capacity, the arriving packets are lost when the queue becomes full. Let Pb_s denote the probability that an arriving packet finds an $\text{MMPP}_s/\text{M}/1/K_s$ queue full (Pb_s will be given by Eq. (7.10)). Therefore, the effective traffic entering the queue of stations is a fraction, $(1-Pb_s)$, of the traffic generated by the station. Given that the splitting of an MMPP give rise to a new MMPP, let MMPP_s^e represent the effective traffic entering the queue of stations.

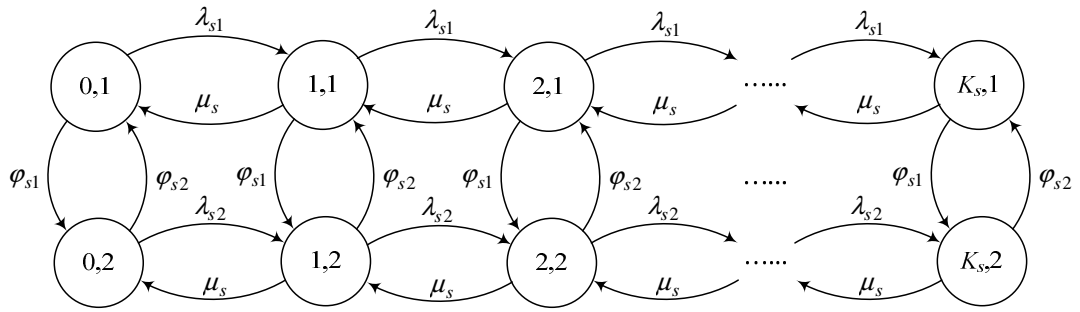

Figure 7.4: State transition rate diagram of $\text{MMPP}_s / \text{M} / 1 / K_s$ queue

MMPP_s^e can be obtained by the splitting of MMPP_s with the splitting probability, $(1 - Pb_s)$. Based on Eq. (3.18), the infinitesimal generator, $\mathbf{Q}_s^e$, and rate matrix, $\mathbf{\Lambda}_s^e$, of MMPP_s^e can be given by

$$\mathbf{Q}_s^e = \mathbf{Q}_s \quad \text{and} \quad \mathbf{\Lambda}_s^e = (1 - Pb_s)\mathbf{\Lambda}_s \tag{7.4}$$

Therefore, the throughput at the station is the mean arrival rate, $\overline{\lambda}_s^e$, of MMPP_s^e. $\overline{\lambda}_s^e$ can be obtained by virtue of Eq. (2.11).

In what follows, the packet delay at the stations will be derived. Let $P_{v,j}$, $(0 \le v \le K_s)$ and $(j = 1, 2)$, represent the probability that there are v packets in the queue of stations and the underlying Markov chain of the MMPP_s, characterising the traffic arriving at the queue of stations, is at state j can be obtained by a bi-variate Markov chain, as shown in Fig. 7.4. State (v, j) corresponds to the case that there are v packets in the queue and the MMPP_s is at state j. The transition rate out of state (v, j) to $(v+1, j)$ is λ_{s_j}, where λ_{s_j} is the arrival rate to the queue of stations and can be obtained from $\mathbf{\Lambda}_s$, while the rate from $(v+1, j)$ to (v, j) is μ_s. The transition rate out of state $(v, 1)$ to $(v, 2)$ is φ_{s1}, while the rate from $(v, 2)$ to $(v, 1)$ is φ_{s2}, where φ_{s1} and φ_{s2} can be given by $\mathbf{Q}_s$.

The transition rate matrix, $\mathbf{G}$, of the bi-variate Markov chain can be readily obtained from Fig. 7.4. The steady-state probability vector, $\mathbf{P} = (P_{v,j}) = (\mathbf{P}_0, \mathbf{P}_1, \cdots, \mathbf{P}_{K_s})$, where $\mathbf{P}_v = (P_{v,1}, P_{v,2})$, $0 \le v \le K_s$, satisfies the following equations

$$\mathbf{PG} = 0 \text{ and } \mathbf{Pe} = 1 \tag{7.5}$$

Solving these equations yields the steady-state vector as [31]

$$\mathbf{P} = \mathbf{u}(\mathbf{I} - \mathbf{\Phi} + \mathbf{eu})^{-1} \tag{7.6}$$

where $\mathbf{\Phi} = \mathbf{I} + \mathbf{G} / \min\{\mathbf{G}(i,i)\}$, $\mathbf{u}$ is an arbitrary row vector of $\mathbf{\Phi}$, and $\mathbf{e}$ is a unit column vector.

Let P_v, $0 \le v \le K_s$, denote the probability that there are v packets in the queue of stations. P_v can be given by

$$P_v = \sum_{j=1}^{2} P_{v,j} \tag{7.7}$$

The delay experienced by a packet to go through a station is equal to the packet waiting time in the queueing system. Let D_s denote the delay that a packet experiences at the station. According to Little's Law [38], we have

$$D_s = \frac{\sum_{v=0}^{K_s} v P_v}{\overline{\lambda}_s^e} \tag{7.8}$$

To determine Pb_s used in Eq. (7.4), let us first calculate the probability, P_v', $0 \le v \le K_s$, that there are v packets in an $\text{MMPP}_s / \text{M} / 1 / K_s$ queue seen by an arriving packet. P_v' can be given by [78]

$$P_v' = \left(\sum_{v=0}^{K_c} \mathbf{P}_v \mathbf{\Lambda}_s \mathbf{e} \right)^{-1} \mathbf{P}_v \mathbf{\Lambda}_s \mathbf{e} \tag{7.9}$$

Therefore, the probability, Pb_s, that an arriving packet finds the finite buffer full can be written as

$$Pb_s = P_{K_s}' \tag{7.10}$$

7.2.3 Packet delay and throughput at access points

Packets departing from the station need to enter the queue of its associated AP. To obtain the arrival process of the traffic arriving at the queue of the AP, the output process from the queue of the station needs to be derived. The output process of the $\text{MMPP}_s / \text{M} / 1 / K_s$ finite buffer queueing system is treated as the output process of $\text{MMPP}_s^e / \text{M} / 1$ infinite buffer queueing system. Let MMPP_s^o represent the output process from the queue of stations. Following the method used in Chapter 6.2.2 to derive the output process from an $\text{MMPP} / \text{M} / 1$ queue, the infinitesimal generator, $\mathbf{Q}_s^o$, and rate matrix, $\mathbf{\Lambda}_s^o$, of MMPP_s^o can be obtained.

The traffic arriving at the queue in the AP can be obtained by applying the flow conservation law [38], i.e., the total departures of packets from all stations in a WLAN must be equal to the total arrivals of packets to the queue of their associated AP. Let MMPP_a denote the traffic arriving at the queue of the AP. Since there are N stations in a WLAN, MMPP_a is the superposition of N traffic flows modelled by MMPP_s^o. Based on Eqs. (3.17)-(3.34), the infinitesimal generator, $\mathbf{Q}_a$, and rate matrix, $\mathbf{\Lambda}_a$, of MMPP_a can be determined. The effective traffic entering the queue

of APs, denoted by MMPP_a^e, can be determined based on Eq. (7.4). The throughput at the AP is the mean arrival rate, $\overline{\lambda}_a^e$, of MMPP_a^e, and can be obtained based on Eq. (2.11). To calculate the packet delay at the APs, the AP is modelled as an $\text{MMPP}/\text{M}/1/K$ queueing system, where the arrival process is MMPP_a and the service rate, μ_a, can be determined by virtue of Eqs. (6.5)-(6.7). According to Eqs. (7.5)-(7.8), the packet delay, D_a, at the APs can be determined. Based on Eqs. (6.13) and (6.14), the output process, MMPP_a^o, from the queue of the APs can be obtained.

7.2.4 Packet delay and throughput at mesh routers

Packets departing from the queue of the APs needs to enter the queue of their associated MR or get lost when the buffer of the MR is full. Let MMPP_{r_i}, $\text{MMPP}_{r_i}^e$, and $\text{MMPP}_{r_i}^o$ denote the traffic arriving at, entering, and departing from the queue of MRs located at i hops away from the gateway, where $1 \le i \le h_{max}$. Recall that MRs perform the dual tasks of both forwarding packets and providing network access to the stations. Since MRs located at the rim of the network do not relay packets from MRs that are located farther away, the packet arriving at the queue of an MR located at the rim of the network is the packet departing from its associated AP, i.e., MMPP_a^o. According to Eq. (7.4), we can obtain the parameter matrices of $\text{MMPP}_{r_{h_{max}}}^e$, characterising the traffic entering the MRs located at the rim of the network. Based on Eqs. (6.13) and (6.14), the output process, $\text{MMPP}_{r_{h_{max}}}^o$, from the MRs that are located at the rim of the network can be determined.

The traffic arriving at the queue of MRs located at i, $(1 \le i \le h_{max} - 1)$, hops

away from the gateway is the superposition of the traffic departing from MRs located at $(i+1)$ hops away and the traffic departing from the associated AP. Since there are N_{i+1} MRs located at $(i+1)$ hops away from the gateway, the traffic departing from the queue of these MRs, denoted by $\text{MMPP}^{o}_{tr_{i+1}}$, is the superposition of N_{i+1} traffic flows modelled by $\text{MMPP}^{o}_{r_{i+1}}$, where $\text{MMPP}^{o}_{r_{i+1}}$ can be obtained based on the derivation of $\text{MMPP}^{o}_{r_{h_{max}}}$. Based on Eqs. (3.17)-(3.34), the parameter matrices of $\text{MMPP}^{o}_{tr_{i+1}}$ can be obtained. Given that the traffic departing from the MRs located at $(i+1)$ hops away from the gateway will be evenly distributed on the MRs located at i hops away, the traffic arriving at an MR located at i hops away from the gateway fed by the traffic departing from the MRs located farther away, denoted by MMPP'_{r_i}, is the splitting of the traffic flow, $\text{MMPP}^{o}_{tr_{i+1}}$, with the splitting probability, $1/N_i$. Based on Eq. (3.18), the infinitesimal generator, $\mathbf{Q}'_{r_i}$, and rate matrix, $\mathbf{\Lambda}'_{r_i}$, of MMPP'_{r_i} can be obtained.

The traffic arriving at the queue of an MR located at i, $(1 \le i \le h_{max}-1)$, hops away from the gateway is the superposition of one MMPP'_{r_i} and one MMPP^{o}_{a}. Following Eqs. (3.17)-(3.34), the infinitesimal generator, $\mathbf{Q}_{r_i}$, and rate matrix, $\mathbf{\Lambda}_{r_i}$, of MMPP_{r_i} can be determined.

The parameter matrices of $\text{MMPP}^{e}_{r_i}$ and $\text{MMPP}^{o}_{r_i}$ can be determined based on the derivation of $\text{MMPP}^{e}_{r_{h_{max}}}$ and $\text{MMPP}^{o}_{r_{h_{max}}}$, respectively. The throughput at the MRs located at i hops away from the gateway is the mean arrival rate, $\overline{\lambda}^{e}_{r_i}$, of

$\text{MMPP}_{r_i}^e$, and can be determined based on Eq. (2.11). The delay for a packet to go through an MR located at i hops away from the gateway, D_{r_i}, can be obtained according to Eqs. (7.5)-(7.8).

7.2.5 Location-dependent end-to-end delay and throughput

The end-to-end delay, D_i^{e2e}, $(1 \le i \le h_{max})$, experienced by a packet from its source to the gateway is equal to the sum of the packet transmission time, the delays at the source station, its associated AP, and all intermediate MRs, respectively. Therefore, we have

$$D_i^{e2e} = i \times t_m + 2 \times t_\ell + D_s + D_a + \sum_{j=1}^{i} D_{r_j} \tag{7.11}$$

The end-to-end throughput, T_i, $(1 \le i \le h_{max})$, is defined as the average number of packets generated by the station in a WLAN, which is associated with the MR located at i hops away from the gateway, successfully received by the gateway. For the purpose of illustration, let us denote such a station as the *i*-hop station. T_i can be determined by the arrival rate of packets generated by an *i*-hop station and these packets are not blocked by its source, associated AP and the intermediate MRs. Therefore, T_i is equal to the product of the arrival rate of the queue in the station and the end-to-end non-blocking probability (i.e., the product of non-blocking probabilities at the station, APs, and all intermediate MRs) as

$$T_i = \overline{\lambda}_s \times (1 - Pb_s) \times (1 - Pb_a) \times \prod_{j=1}^{i} (1 - Pb_{r_i}) \tag{7.12}$$

where $\overline{\lambda}_s$ is the mean arrival rate of MMPP_s and can be determined based on Eq.

(2.11); Pb_a and Pb_{r_i} can be obtained by virtue of Eq. (7.10).

7.2.6 QoS performance metrics

In this section, the aggregate network throughput and average end-to-end delay experienced by the packet are derived as the QoS performance metrics of the integrated WLANs and Internet-access mesh networks. The aggregate network throughput, T, is defined as the total number of packets received by the gateway and can be given by

$$T = \sum_{i=1}^{h_{max}} N \times N_i \times T_i \tag{7.13}$$

The average end-to-end delay, D_{e2e}, experienced by the packet in the integrated wireless networks can be obtained by averaging the delays of packets that are successfully received by the gateway. Therefore, we have

$$D_{e2e} = \frac{\sum_{i=1}^{h_{max}} N \times N_i \times T_i \times D_i^{e2e}}{T} = \frac{\sum_{i=1}^{h_{max}} N \times N_i \times T_i \times D_i^{e2e}}{\sum_{i=1}^{h_{max}} N \times N_i \times T_i} \tag{7.14}$$

7.3 Validation of the model

The accuracy of the analytical model is validated through a discrete-event simulator based on the OMNeT++ simulation platform [89]. Each simulation experiment was run until the network reached its steady state, which means that a further increase in the simulation time does not change the statistical performance results significantly, and collected based on Eqs. (3.51) and (3.52). The packets generated by each station

Table 7.2: Successful channel-access probabilities for MRs located at *i* hops away from the gateway in integrated WLANs and Internet-access mesh networks

i	**1**	**2**	**3**	**4**
P_{r_i}	8.2e-2	7.3e-2	6.4e-2	5.5e-2

are modelled by MMPP$_s$ with the infinitesimal generator, $\mathbf{Q}_s$, and rate matrix, $\mathbf{\Lambda}_s$. Extensive simulation experiments have been performed for this purpose. However, for the sake of specific illustration and without loss of generality, the results are presented for the following cases:

- the number of MRs in the WMN: C = 81 MRs;
- the number of stations in a WLAN: N = 3, 5, 7, and 10 stations;
- the channel capacity in the WLANs and in the WMN is 11 Mb/s and 54 Mb/s, respectively;
- the packet size, ω, is set to 8000 and 10000 bits;
- the length of a time slot is set to be the amount of airtime required to transmit one packet, i.e., t_ℓ = 7.27e-4 s and 9.09e-4 in the WLAN and t_m = 1.48e-4 s and 1.85e-4 in the WMN;
- the buffer size is set to 32 packets in the station, 48 packets in the AP, and 64 packets in the MR due to the increase of traffic loads in stations, APs and MRs;
- the successful channel-access probabilities, P_s, P_a, and P_{r_i} $(1 \le i \le h_{max})$, for stations, APs, and MRs are given by P_s = 2.5e-3, P_a = 7e-3, and P_{r_i}, respectively, which are shown in Table 7.2;
- the infinitesimal generator, $\mathbf{Q}_s$, of MMPP$_s$ are set as follows, representing different degrees of traffic burstiness and correlations:

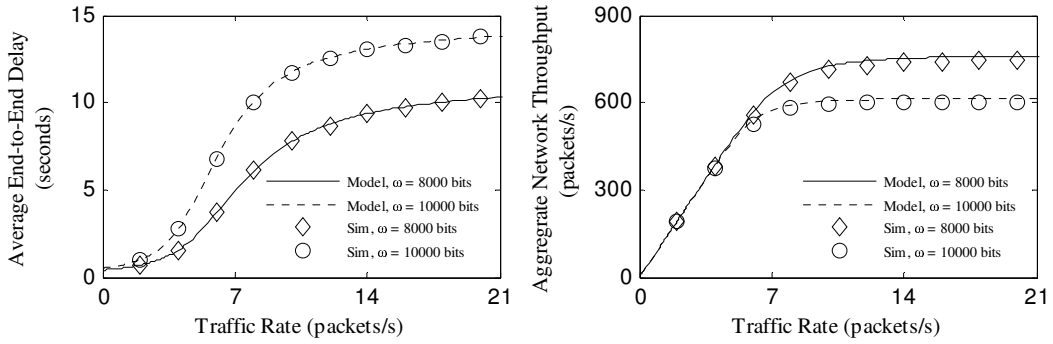

Figure 7.5: Average end-to-end delay and aggregate network throughput predicted by the model and simulation in the integrated WLANs and Internet-access mesh networks with $N = 3$, and $\varphi_{s1} = 0.6$, $\varphi_{s2} = 0.4$

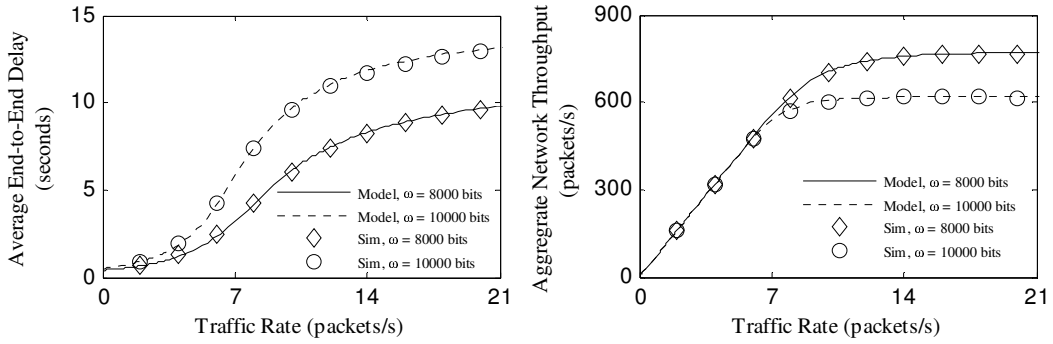

Figure 7.6: Average end-to-end delay and aggregate network throughput predicted by the model and simulation in the integrated WLANs and Internet-access mesh networks with $N = 5$, and $\varphi_{s1} = 0.8$, $\varphi_{s2} = 0.2$

$$\begin{bmatrix} -0.6 & 0.6 \\ 0.4 & -0.4 \end{bmatrix}, \begin{bmatrix} -0.8 & 0.8 \\ 0.2 & -0.2 \end{bmatrix}, \begin{bmatrix} -0.5 & 0.5 \\ 0.1 & -0.1 \end{bmatrix}, \text{ and } \begin{bmatrix} -0.03 & 0.03 \\ 0.02 & -0.02 \end{bmatrix}.$$

Since the average end-to-end delay and aggregate network throughput are the key performance metrics to evaluate the integrated WLANs and Internet-access mesh networks, we focus on the two important metrics to validate the accuracy of the analytical model. Figs. 7.5-7.8 depict the analytical and simulation results for the average end-to-end delay and aggregate network throughput as a function of the

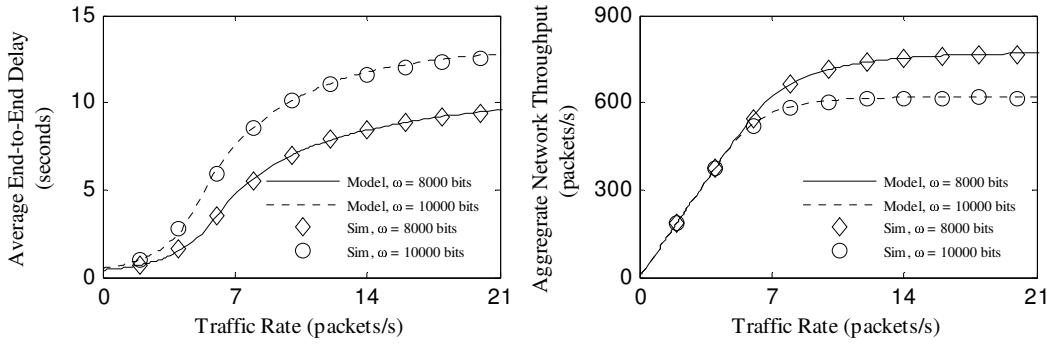

Figure 7.7: Average end-to-end delay and aggregate network throughput predicted by the model and simulation in the integrated WLANs and Internet-access mesh networks with $N = 7,\ \varphi_{s1} = 0.5,\ \varphi_{s2} = 0.1$

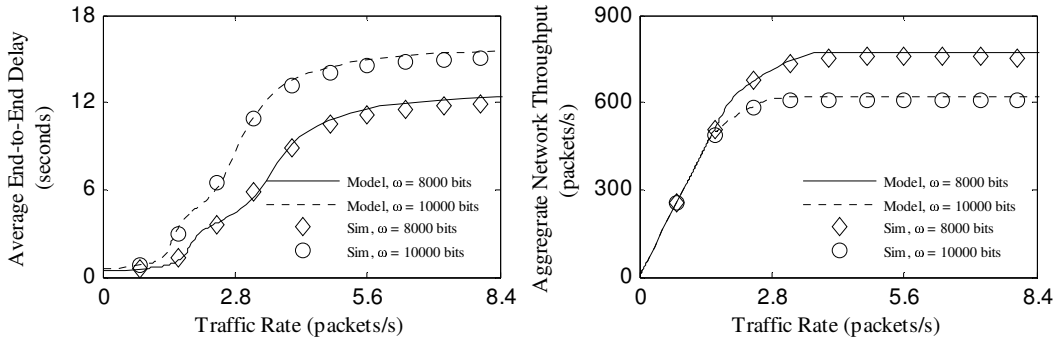

Figure 7.8: Average end-to-end delay and aggregate network throughput predicted by the model and simulation in the integrated WLANs and Internet-access mesh networks with $N = 10,\ \varphi_{s1} = 0.03,\ \varphi_{s2} = 0.02$

traffic rate of the station in the integrated wireless networks. In these figures, the horizontal axis represents the traffic rate, λ_{s1}, at which a station injects messages into the network when the MMPP_s is at state 1, while the vertical axis denotes the performance metrics obtained from the analytical model. For the sake of clarity of the figures, the arrival rate, λ_{s2}, at state 2 is deliberately set to zero; otherwise the three-dimensional graphs will be used to represent the results. These figures reveal that the results obtained from the analytical model closely match those obtained from

the simulation.

7.4 Performance analysis

The performance bottlenecks which can deteriorate the system efficiency often exist in various types of networks, including wired and wireless networks. Removing the bottlenecks is a significant task in the network design and performance analysis. Numerous research efforts have been made to address this problem in wireless networks. For example, Jain, Gupta, and Agrawal [53] investigated various issues involved in the design of the MAC protocol for multi-hop wireless networks and studied the problem of bottleneck nodes. They adopted multiple beam smart antennas in each node to increase the throughput of the bottleneck nodes. However, the studies of bottlenecks in the integrated WLANs and WMN have rarely been carried out and remain a vital problem. In this section, the bottlenecks in the integrated wireless networks will be investigated with the aim of obtaining the important insights to alleviate the bottleneck problem.

From the derivation of MMPP_{r_i} in Chapter 7.2.4, we can find that the traffic is non-uniformly distributed in MRs at different locations with respect to the gateway. The MRs located closer to the gateway receive the higher traffic loads than those farther away. This fact leads to a large fraction of the end-to-end delay experienced by a packet at MRs located closer to the gateway. The forthcoming sub-chapters will present how the analytical model can be used to adjust the system parameters, including successful channel-access probabilities at MRs and the deployment of MRs to achieve a fairly distributed packet delay in MRs at different locations so as to alleviate the bottleneck problem around the gateway.

7.4.1 Effects of successful channel-access probabilities at mesh routers

Since the traffic loads at the MRs located closer to the gateway become higher, the packet delay experienced at these MRs increases. Based on Eq. (6.7), the expression of the service rate of the queue in MR located at i hops away from the gateway can be obtained as follows:

$$\mu_{r_i} \approx \frac{\ln(1-P_{r_i})^{-1}}{t_m} \tag{7.15}$$

The above equation shows that the larger successful channel-access probabilities of MRs can increase the service rate at the queue of these MRs. Therefore, the successful channel-access probabilities of MRs located at i hops away from the gateway, P_{r_i}, can be expressed as

$$P_{r_i} = 1 - e^{-\mu_{r_i} t_m} \tag{7.16}$$

From the derivation of packet delay at MRs, we can find that the service rate can directly affect the packet delay. In other words, increasing the service rate of the queueing system at MRs located closer to the gateway can decrease the packet delay at these MRs. Therefore, tuning the successful channel-access probabilities of MRs at different locations with respect to the gateway can reduce the packet delay and alleviate the bottleneck problem around the gateway.

7.4.2 Effects of deployment of mesh routers

Another approach to controlling the packet delay at MRs closer to the gateway is to adjust the density of MRs located at i hops away from the gateway. If additional

MRs are deployed at the location under heavy traffic, the loads will be shared by more MRs. This strategy can reduce the throughput and packet delay at the MRs in the region. With the additional deployment, the total number of MRs in the network is

$$C + \sum_{i=1}^{h_{max}} A_i \tag{7.17}$$

where A_i is the additional number of MRs to share the traffic at the location with i hops away from the gateway. Let ϕ_i and ϕ_i'' represent the ratio of the traffic arriving at an MR located at i hops and $(i+1)$ hops away from the gateway without and with additional MRs deployed, respectively. ϕ_i and ϕ_i'' can be expressed as

$$\phi_i = \frac{\overline{\lambda}_{r_i}}{\overline{\lambda}_{r_{i+1}}} \quad \text{and} \quad \phi_i'' = \frac{\overline{\lambda}_{r_i}''}{\overline{\lambda}_{r_{i+1}}''} \tag{7.18}$$

where $\overline{\lambda}_{r_i}$ and $\overline{\lambda}_{r_i}''$ are the traffic arrival rates at an MR located at i hops away from the gateway without and with additional MRs deployed. Based on the flow conservation law [38], the total amount of traffic arriving at the MRs located at i hops away from the gateway must be the same before and after the additional MRs are deployed, i.e.,

$$N_i \times \overline{\lambda}_{r_i} = (N_i + A_i) \times \overline{\lambda}_{r_i}'' \tag{7.19}$$

Let $\tilde{A}_i$ denote the total number of MRs located at i hops away from the gateway with additional MRs deployed, i.e., $\tilde{A}_i = N_i + A_i$. The aforementioned equation can be re-written as

$$\overline{\lambda_{r_i}''} = \frac{N_i \times \overline{\lambda}_{r_i}}{\tilde{A}_i} \tag{7.20}$$

Substituting Eq. (7.20) in Eq. (7.18), we have

$$\tilde{A}_i = \frac{\phi_i}{\phi_i''} \times \frac{N_i}{N_{i+1}} \times \tilde{A}_{i+1} \tag{7.21}$$

Therefore, the bottleneck problem around the gateway can be alleviated with $\tilde{A}_i$ additional MRs deployed at the location that is i hops away from the gateway.

7.5 Conclusions

WMNs can provide the last-mile wireless broadband access to the Internet at a low cost. This chapter has presented an analytical model to investigate the QoS performance measures including end-to-end delay and throughput in the integrated WLANs and Internet-access WMN in the presence of bursty traffic. Extensive simulation experiments have been conducted to validate the accuracy of the analytical model and demonstrated that the analytical results predicted by the model closely match those obtained from simulation. The model has been used to investigate the bottleneck problem in the integrated wireless networks. Moreover, by virtue of the developed analytical model, we can tune the system parameters and adjust the density of MRs at different locations with respect to the gateway in order to alleviate this bottleneck problem.

Chapter 8

Conclusions and Future Directions

Traffic patterns have a significant impact on the performance of wired or wireless networks. In this thesis, we have first investigated the performance of wired INs in multi-computer systems by means of analytical modelling in the presence of bursty traffic with non-uniformly distributed message destinations, which can capture the bursty nature of realistic network traffic in the both temporal domain and spatial domain. Secondly, analytical models have been developed to evaluate the performance of INs in heterogeneous multi-cluster systems under bursty traffic with non-uniformly distributed message destinations. Finally, we have proposed analytical models to investigate the performance of integrated WLANs and WMNs under the traffic patterns exhibited by real-world applications. After the validation, the developed analytical models have been adopted as practical and cost-effective tools to investigate the performance of heterogeneous wired or wireless networks. In what follows, we first summarise the results of this research, followed by a discussion of future research directions.

8.1 Summary of the results

The main contributions of this thesis are summarised as follows:

- in Chapter 3, an analytical model has been developed to investigate the performance of INs in multi-computer systems in the presence of bursty traffic with hot-spot destinations. The comparison between the analytical results and those obtained from extensive simulation experiments has shown that the derived analytical model possesses a good degree of accuracy for predicting the network performance under different design alternatives and various traffic conditions. The analytical results have revealed that the network performance degrades considerably under this type of traffic pattern;
- in Chapter 4, an analytical model has been proposed to investigate the performance of INs in multi-cluster systems in the presence of bursty traffic with communication locality. The analytical model has been validated by extensive simulation experiments. The results have shown that the performance measures predicted by the model closely matches those obtained from the simulation. The analytical results have revealed that the bursty traffic can degrade the network performance considerably. Moreover, the results have demonstrated that the communication locality can decrease the degrading effects of bursty traffic on the performance of wired INs in multi-cluster systems;
- in Chapter 5, an analytical model has been developed to investigate the performance of INs in heterogeneous multi-cluster systems under bursty traffic. The model has been able to characterise the heterogeneity in the number of nodes and the configuration of intra-communication networks within each cluster. Moreover, the model has considered a multi-user environment in multi-cluster systems. The comparison between the analytical results and those obtained from extensive simulation experiments has shown that the derived

analytical model possesses a good degree of accuracy for predicting the network performance under different design alternatives and various traffic conditions. The analytical results have revealed the importance of balancing the size of each cluster and the number of clusters in order to maximise the system performance;

- in Chapter 6, an analytical model has been developed to investigate the end-to-end delay in a WMN interconnecting multiple WLANs in the presence of bursty traffic with communication locality. Extensive simulation experiments have been conducted to validate the accuracy of the analytical model. The results obtained from the simulation experiments have confirmed that the derived analytical model exhibits a considerable degree of accuracy. The proposed analytical model has been used to obtain the maximum achievable throughput which reflects the capacity of wireless networks;
- in Chapter 7, an analytical model has been proposed to investigate the QoS performance measures including end-to-end delay and throughput in the integrated WLANs and Internet-access mesh networks in the presence of bursty traffic. Extensive simulation experiments have been conducted to validate the accuracy of the analytical model and demonstrated that the analytical results predicted by the model closely match those obtained from the simulation. The model has been used to investigate the bottleneck problem in integrated wireless networks. Moreover, by virtue of the developed analytical model, we can tune the system parameters and adjust the density of MRs at different locations with respect to the gateway in order to alleviate this bottleneck problem.

8.2 Directions of the future work

In this section, we discuss the open issues and problems in wired or wireless

networks that require further investigations. A selection of such problems is outlined as follows.

- The wired or wireless networks consist of multiple network components. The failure rates of network components increase as the network size increases. Therefore, incorporation of fault-tolerant techniques in these networks is of great importance to ensure the network to keep running in a degraded mode. Several studies have investigated the fault-tolerant routing in wired networks [27] or wireless networks [18] by means of simulation. There is hardly any analytical model of fault-tolerant routing in wired or wireless networks when the network is subject to the traffic patterns exhibited by real-world applications.
- An optical network enables broadband services but is too costly to act as an access network. On the other hand, a wireless network can be easily deployed with low cost but provides a highly bandwidth-constrained transmission channel. In light of the complementary characteristics of optical and wireless technologies, an integration of optical fibre networks and wireless networks is a promising solution to provide a broadband and ubiquitous last-mile connection [35]. It is beneficial to develop analytical models to investigate the performance of wireless-optical networks that use WMNs at the front-end and optical backhaul networks at the back-end for wireless-optical broadband access.
- WMNs have been a key technology to interoperate with diverse wireless systems and provide last-mile wireless broadband connectivity for the Internet access at a low cost. However, the dual usage of wireless communications for both forwarding packets and providing network access to the users makes them very resource dependent and places additional strain on the scarce resources [4]. The cognitive radio has been recognised as a promising and powerful solution to

alleviate the spectrum scarcity that wireless communications face currently [3, 126]. Most existing studies on cognitive mesh networks are mainly focused on the topology management [12], resource allocation [48], and power estimation [13]. The main problem in the literature is the lack of researches regarding the delay and throughput analysis in cognitive mesh networks, in particular when the network is subject to the bursty traffic.

References

[1] J.H. Abawajy, "An Efficient Adaptive Scheduling Policy for High-Performance Computing," *Future Generation Computer Systems*, vol. 25, no. 3, pp. 364-370, 2009.

[2] A. Agarwal, "Limits on Interconnection Network Performance," *IEEE Trans. on Parallel Distributed Systems*, vol. 2, no. 4, pp. 398-412, 1991.

[3] I.F. Akyildiz, W.-Y. Lee, and K.R. Chowdhury, "CRAHNs: Cognitive Radio Ad Hoc Networks," *Ad Hoc Networks*, vol. 7, no. 5, pp. 810-836, 2009.

[4] I.F. Akyildiz and X. Wang, "A Survey on Wireless Mesh Networks," *IEEE Communications Magazine*, vol. 43, no. 9, pp. S23-S30, 2005.

[5] N. Alzeidi, M. Ould-Khaoua, and A. Khonsari, "A New General Method to Compute Virtual Channels Occupancy Probabilities in Wormhole Networks," *Journal of Computer and System Sciences*, vol. 74, no. 6, pp. 1033-1042, 2008.

[6] G. Athanasiou, T. Korakis, O. Ercetin, and L. Tassiulas, "A Cross-Layer Framework for Association Control in Wireless Mesh Networks," *IEEE Trans. on Mobile Computing*, vol. 8, no. 1, pp. 65-80, 2009.

[7] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535-547, 2000.

[8] N. Bisnik and A. Abouzeid, "Delay and Throughput in Random Access

Wireless Mesh Networks," in *Proc. of IEEE International Conference on Communications (ICC'06)*, pp. 403-408, 2006.

[9] V. Brik, S. Rayanchu, S. Saha, S. Sen, V. Shrivastav, and S. Banerjee, "A Measurement Study of A Commercial-Grade Urban WiFi Mesh," in *Proc. of ACM SIGCOMM Conference on Internet Measurement (IMC'08)*, pp. 111-124, 2008.

[10] L. Cai, X. Shen, and J.W. Mark, *Multimedia Services in Wireless Internet: Modeling and Analysis*, John Wiley & Sons Ltd, 2009.

[11] M. Cao, W. Ma, Q. Zhang, and X. Wang, "Analysis of IEEE 802.16 Mesh Mode Scheduler Performance," *IEEE Trans. on Wireless Communications*, vol. 6, no. 4, pp. 1455-1464, 2007.

[12] T. Chen, H. Zhang, G.M. Maggio, and I. Chlamtac, "Topology Management in CogMesh: A Cluster-Based Cognitive Radio Mesh Network," in *Proc. of IEEE International Conference on Communications (ICC'07)*, pp. 6516-6521, 2007.

[13] K.R. Chowdhury and I.F. Akyildiz, "Cognitive Wireless Mesh Networks with Dynamic Spectrum Access," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 168-181, 2008.

[14] A. Clematis and A. Corana, "Modeling Performance of Heterogeneous Parallel Computing Systems," *Parallel Computing*, vol. 25, no. 9, pp. 1131-1145, 1999.

[15] F. Dai and J. Wu, "Proactive Route Maintenance in Wireless Ad Hoc Networks," in *Proc. of IEEE International Conference on Communications (ICC'05)*, vol. 2, pp. 1236-1240, 2005.

[16] W.J. Dally, "Virtual Channel Flow Control," *IEEE Trans. on Parallel and*

Distributed Systems, vol. 3, no. 2, pp. 194-205, 1992.

[17] W.J. Dally and B.P. Towles, *Principles and Practices of Interconnection Network*, Morgan Kaufmann, 2004.

[18] A. Datta, "A Fault-tolerant Protocol for Energy-efficient Permutation Routing in Wireless Networks " *IEEE Trans. on Computers*, vol. 54, no. 11, pp. 1409-1421, 2005.

[19] P.A. Dinda, "Design, Implementation, and Performance of an Extensible Toolkit for Resource Prediction in Distributed Systems," *IEEE Trans. on Parallel and Distributed Systems*, vol. 17, no. 2, pp. 160-173, 2006.

[20] J. Dongarra and A. Lastovetsky, "An Overview of Heterogeneous High Performance and Grid Computing," *Engineering the Grid: Status and Perspective*, ISBN: 1-58883-038-1, 2006.

[21] J.T. Draper and J. Ghosh, "A Comprehensive Analytical Model for Wormhole Routing in Multicomputer Systems," *Journal of Parallel & Distributed Computing*, vol. 32, no. 2, pp. 202-214, 1994.

[22] X. Du, X. Zhang, and Z. Zhu, "Memory Hierarchy Considerations for Cost-Effective Cluster Computing," *IEEE Trans. on Computers*, vol. 49, no. 9, pp. 915-933, 2000.

[23] J. Duato, S. Yalamanchili, and L. Ni, *Interconnection Networks: An Engineering Approach*, Morgan Kaufmann, 2003.

[24] I. Elhanany and M. Kahane, "Heterogeneous Bursty Traffic Dispersion over Multiple Server Clusters," *IEEE Communications Letters*, vol. 9, no. 3, pp. 261-263, 2005.

[25] S.M. Faccin, C. Wijting, J. Kenckt, and A. Damle, "Mesh WLAN Networks: Concept and System Design," *IEEE Wireless Communications*, vol. 13, no. 2,

pp. 10-17, 2006.

[26] A. Fallahi, E. Hossain, and A.S. Alfa, "QoS and Energy Trade Off in Distributed Energy-Limited Mesh/Relay Networks: A Queuing Analysis," *IEEE Trans. on Parallel and Distributed Systems*, vol. 17, no. 6, pp. 576-592, 2006.

[27] M.H. Farahabady, F. Safaei, A. Khonsari, and M. Fathy, "Characterization of Spatial Fault Patterns in Interconnection Networks," *Parallel Computing*, vol. 32, no. 11-12, pp. 886-901, 2006.

[28] A. Farbod and B. Liang, "Structured Admission Control Policy in Heterogeneous Wireless Networks with Mesh Underlay," in *Proc. of IEEE Conference on Computer Communications (INFOCOM'09)*, pp. 495-503, 2009.

[29] H.-W. Ferng and J.-F. Chang, "Connection-Wise End-to-End Performance Analysis of Queuing Networks with MMPP Inputs," *Performance Evaluation*, vol. 43, no. 1, pp. 39-62, 2001.

[30] H.-W. Ferng and J.-F. Chang, "Departure Processes of BMAP/G/1 Queues," *Queueing Systems: Theory and Applications*, vol. 39, no. 2-3, pp. 109-135, 2001.

[31] W. Fischer and K. Meier-Hellstern, "The Markov-Modulated Poisson Process (MMPP) Cookbook," *Performance Evaluation*, vol. 18, no. 2, pp. 149-171, 1993.

[32] F. Frattolillo, "Supporting Data Management on Cluster Grids," *Future Generation Computer Systems*, vol. 24, no. 2, pp. 166-176, 2008.

[33] A.E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-Delay Trade-Off in Wireless Networks," in *Proc. of IEEE International Conference*

on Computer Communications (INFOCOM'04), pp. 464-475, 2004.

[34] M. Garetto, T. Salonidis, and E.W. Knightly, "Modeling Per-Flow Throughput and Capturing Satrvation in CSMA Multi-Hop Wireless Networks," *IEEE/ACM Trans. on Networking*, vol. 16, no. 4, pp. 864-877, 2008.

[35] N. Ghazisaidi, M. Maier, and C.M. Assi, "Fiber-Wireless (FiWi) Access Networks: A Survey," *IEEE Communications Magazine*, vol. 47, no. 2, pp. 160-167, 2009.

[36] C. Gomez, F. Gilabert, M.E. Gomez, P. Lopez, and J. Duato, "Deterministic Versus Adaptive Routing in Fat-Trees," in *Proc. of IEEE International Conference on Parallel and Distributed Processing Symposium (IPDPS'07)*, pp. 1-8, 2007.

[37] C. Gomez, M.E. Gomez, P. Lopez, and J. Duato, "FT^{2}EI: A Dynamic Fault-Tolerant Routing Methodology for Fat Trees with Exclusion Intervals," *IEEE Trans. on Parallel and Distributed Systems*, vol. 20, no. 6, pp. 802-817, 2009.

[38] D. Gross and C. Harris, *Fundamentals of Queueing Theory*, 3rd ed, Hoboken, NJ, Wiley, 1998.

[39] M. Guizani, P. Lin, S.-M. Cheng, and D.-W. Huang, "Performance Evaluation for Minislot Allocation for Wireless Mesh Networks," *IEEE Trans. Vehicular Technology*, vol. 57, no. 6, pp. 3732-3745, 2008.

[40] J. Ha and C.-H. Choi, "TCP Fairness for Uplink and Downlink Flows in WLANs," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM'06)*, pp. 1-5, 2006.

[41] H. Hashemi-Najafabadi, H. Sarbazi-Azad, and P. Rajabzadeh, "An Accurate Performance Model of Fully Adaptive Routing in Wormhole-Switched Two-Dimensional Mesh Multicomputers," *Microprocessors and Microsystems*, vol.

31, no. 7, pp. 445-455, 2007.

[42] R. He and J.G. Delgado-Frias, "Fault Tolerant Interleaved Switching Fabrics for Scalable High-Performance Routers," *IEEE Trans. on Parallel and Distributed Systems*, vol. 18, no. 12, pp. 1727-1739, 2007.

[43] H. Heffes, "A Class of Data Traffic Processes-Covariance Function Characterization and Related Queueing Results," *Bell System Technical Journal*, vol. 59, no. 6, pp. 897-929, 1980.

[44] H. Heffes and D.M. Lucantoni, "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance," *IEEE Journal on Selected Areas in Communications*, vol. 4, no. 6, pp. 856-867, 1986.

[45] A. Heindl, "Decomposition of General Queueing Networks with MMPP Inputs and Customer Losses," *Performance Evaluation*, vol. 51, no. 2-4, pp. 117-136, 2003.

[46] G.R. Hiertz, Y. Zang, S. Max, T. Junge, E. Weiss, and B. Wolz, "IEEE 802.11s: WLAN Mesh Standardization and High Performance Extensions," *IEEE Network*, vol. 22, no. 3, pp. 12-19, 2008.

[47] E. Hossain, *Heterogeneous Wireless Access Networks: Architectures and Protocols*, Springer, 2008.

[48] Y.T. Hou, Y. Shi, and H.D. Sherali, "Spectrum Sharing for Multi-Hop Networking with Cognitive Radios," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 146-155, 2008.

[49] F. Houéto, S. Pierre, R. Beaubrun, and Y. Lemieux, "Reliability and Cost Evaluation of Third-Generation Wireless Access Network Topologies: A Case Study," *IEEE Trans. on Reliability*, vol. 51, no. 2, pp. 229-239, 2002.

[50] J. Hu, G. Min, and M.E. Woodward, "Analysis and Comparison of Burst Transmission Schemes in Unsaturated 802.11e WLANs," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM'07)*, pp. 5133-5137, 2007.

[51] C.-L. Huang and W. Liao, "Throughput and Delay Performance of IEEE 802.11e Enhanced Distributed Channel Access (EDCA) Under Saturation Condition," *IEEE Trans. on Wireless Communications*, vol. 6, no. 1, pp. 136-145, 2007.

[52] J.-H. Huang, L.-C. Wang, and C.-J. Chang, "Capacity and QoS for a Scalable Ring-Based Wireless Mesh Network," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2070-2080, 2006.

[53] V. Jain, A. Gupta, and D.P. Agrawal, "On-Demand Medium Access in Multihop Wireless Networks with Multiple Beam Smart Antennas," *IEEE Trans. on Parallel and Distributed Systems*, vol. 19, no. 4, pp. 489-502, 2008.

[54] B. Javadi, J.H. Abawajy, and M.K. Akbari, "A Comprehensive Analytical Model of Interconnection Networks in Large-Scale Cluster Systems," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 1, pp. 75-97, 2007.

[55] B. Javadi, M.K. Akbari, and J.H. Abawajy, "A Performance Model for Analysis of Heterogeneous Multi-Cluster Systems," *Parallel Computing*, vol. 32, no. 11-12, pp. 831-851, 2006.

[56] B. Javadi, M.K. Akbari, and J.H. Abawajy, "Multi-Cluster Computing Interconnection Network Performance Modeling and Analysis," *Future Generation Computer Systems*, vol. 25, no. 7, pp. 737-746, 2009.

[57] H.-J. Ju and I. Rubin, "Mesh Topology Construction for Interconnected

Wireless LANs," in *Proc. of Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON'05)*, pp. 284-294, 2005.

[58] Z. Juhasz, P. Kacsuk, and D. Kranzlmuller, *Distributed and Parallel Systems: Cluster and Grid Computing*, Springer, 2004.

[59] A. Khonsari, H. Sarbazi-Azad, and M. Ould-Khaoua, "An Analytical Model of Adaptive Wormhole Routing with Time-out," *Future Generation Computer Systems*, vol. 19, no. 1, pp. 1-12, 2003.

[60] J. Kim and C.R. Das, "Hypercube Communication Delay with Wormhole Routing," *IEEE Trans. on Computers*, vol. 43, no. 7, pp. 806-814, 1994.

[61] J.S. Kim and D.J. Lilja, "Characterization of Communication Patterns in Message-Passing Parallel Scientific Application Programs," in *Proc. of 2nd International Workshop on Communication, Architecture, and Applications for Network-Based Parallel Computing (CANPC'98)*, Lecture Notes in Computer Science, vol. 1362, Springer, pp. 202-216, 1998.

[62] L. Kleinrock, *Queueing Systems*, vol. 1, New York, John Wiley, 1975.

[63] K. Kotapati and S.P. Dandamudi, "Buffer Management in Wormhole-Routed Torus Multicomputer Networks," *Future Generation Computer Systems*, vol. 16, no. 5, pp. 483-491, 2000.

[64] D.D. Kouvatsos, S. Assi, and M. Ould-khaoua, "Performance Modelling of Wormhole- Routed Hypercubes with Bursty Traffic and Finite Buffers," *International Journal of Simulation, Practics, Systems, Science and Technology*, vol. 6, no. 3-4, pp. 69-81, 2005.

[65] D.B. Larkins, J. Dinan, S. Krishnamoorthy, S. Parthasarathy, A. Rountev, and P. Sadayappan, "Global Trees: A Framework for Linked Data Structures on

Distributed Memory Parallel Systems," in *Proc. of ACM/IEEE Conference on Supercomputing*, CD-ROM, 2008.

[66] L. Le and E. Hossain, "Tandem Queue Models with Applications to QoS Routing in Multihop Wireless Networks," *IEEE Trans. on Mobile Computing*, vol. 7, no. 8, pp. 1025-1040, 2008.

[67] L.B. Le, A.T. Nguyen, and E. Hossain, "A Tandem Queue Model for Performance Analysis in Multihop Wireless Networks," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC'07)*, pp. 2981-2985, 2007.

[68] J. Li, C. Blake, D.S.J. De, C. Hu, I. Lee, and R. Morris, "Capacity of Ad Hoc Wireless Networks," in *Proc. of ACM International Conference on Mobile Computing and Networking (MOBICOM'01)*, pp. 61-69, 2001.

[69] X.-Y. Lin, Y.-C. Chung, and T.-Y. Huang, "A Multiple LID Routing Scheme for Fat-Tree-Based InfiniBand Networks," in *Proc. of IEEE International Parallel and Distributed Processing Symposium (IPDPS'04)*, CD-ROM, 2004.

[70] K.-H. Liu, X. Ling, X. Shen, and J.W. Mark, "Performance Analysis of Prioritized MAC in UWB WPAN With Bursty Multimedia Traffic," *IEEE Trans. Vehicular Technology*, vol. 57, no. 4, pp. 2462-2473, 2008.

[71] T. Liu and W. Liao, "Location-Dependent Throughput and Delay in Wireless Mesh Networks," *IEEE Trans. Vehicular Technology*, vol. 57, no. 2, pp. 1188-1198, 2008.

[72] S. Loucif and M. Ould-Khaoua, "The Impact of Virtual Channel Allocation on the Performance of Deterministic Wormhole-Routed *k*-Ary *n*-Cubes," *Simulation Modelling Practice and Theory*, vol. 10, no. 8, pp. 525-541, 2002.

[73] S. Loucif, M. Ould-Khaoua, and L.M. Mackenzie, "Analysis of Fully

Adaptive Wormhole Routing in Tori," *Parallel Computing*, vol. 25, no. 12, pp. 1477-1487, 1999.

[74] S. Loucif, M. Ould-Khaoua, and G. Min, "A Queueing Model for Predicting Message Latency in Uni-Directional *k*-Ary *n*-Cubes with Deterministic Routing and Non-Uniform Traffic," *Cluster Computing*, vol. 10, no. 2, pp. 229-239, 2007.

[75] R. Luo, D. Belis, R.M. Edwards, and G.A. Manson, "Estimation of Average Hop Count Using the Grid Pattern in Multi-Hop Wireless Ad-Hoc Network," in *Proc. of London Communication Symposium*, pp. 13-16, 2002.

[76] D. Malone, K. Duffy, and D. Leith, "Modeling the 802.11 Distributed Coordination Function in Nonsaturated Heterogeneous Conditions," *IEEE/ACM Trans. on Networking*, vol. 15, no. 1, pp. 159-172, 2007.

[77] R. Marculescu, U.Y. Ogras, L.-S. Peh, N.E. Jerger, and Y. Hoskote, "Outstanding Research Problems in NoC Design: System, Microarchitecture, and Circuit Perspectives," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 28, no. 1, pp. 3-21, 2009.

[78] K.S. Meier-Hellstern, "The Analysis of A Queue Arising in Overflow Models," *IEEE Trans. on Communications*, vol. 37, no. 4, pp. 367-372, 1989.

[79] J. Miguel-Alonso, C. Izu, and J.A. Gregorio, "Improving the Performance of Large Interconnection Networks Using Congestion-Control Mechanisms," *Performance Evaluation*, vol. 65, no. 3-4, pp. 203-211, 2008.

[80] G. Min, J. Hu, and M.E. Woodward, "An Analytical Model of the TXOP Scheme with Heterogeneous Classes of Stations," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM'08)*, pp. 1-5, 2008.

[81] G. Min and M. Ould-Khaoua, "A Performance Model for Wormhole-Switched

Interconnection Networks under Self-Similar Traffic," *IEEE Trans. on Computers*, vol. 53, no. 5, pp. 601-613, 2004.

[82] G. Min and M. Ould-Khaoua, "Performance Modelling and Evaluation of Virtual Channels in Multicomputer Networks with Bursty Traffic," *Performance Evaluation*, vol. 58, no. 2-3, pp. 143-162, 2004.

[83] G. Min, Y. Wu, K. Li, and A.Y. Al-Dubai, "Performance Modelling and Optimization of Integrated Wireless LANs and Multi-Hop Mesh Networks," *International Journal of Communication Systems*, 2009.

[84] G. Min, Y. Wu, K. Li, and B. Javadi, "A Performance Model for Multi-Cluster Computing Systems in the Presence of Hot-Spot Traffic," in *Proc. of IEEE International Conference on Communications (ICC'10)*, 2010.

[85] G. Min, Y. Wu, M. Ould-Khaoua, H. Yin, and K. Li, "Performance Modelling and Analysis of Interconnection Networks with Spatio-Temporal Bursty Traffic," in *Proc. of Global Communications Conference (GLOBECOM'09)*, 2009.

[86] M.J. Neely, "Delay Analysis for Maximal Scheduling With Flow Control in Wireless Networks With Bursty Traffic," *IEEE/ACM Trans. on Networking*, vol. 17, no. 4, pp. 1146-1159, 2009.

[87] D. Niyato and E. Hossain, "Integration of IEEE 802.11 WLANs with IEEE 802.16-based Multihop Infrastructure Mesh/Relay Networks: A Game-Theoretic Approach to Radio Resource Management," *IEEE Network*, vol. 21, no. 3, pp. 6-14, 2007.

[88] D. Niyato and E. Hossain, "Dynamics of Network Selection in Heterogeneous Wireless Networks: An Evolutionary Game Approach," *IEEE Trans. on Vehicular Technology*, vol. 58, no. 4, pp. 2008-2017, 2009.

[89] OMNeT++ Network Simulator, http://www.omnetpp.org/

[90] J.M. Orduna and J. Duato, "A High Performance Router Architecture for Multimedia Applications," in *Proc. of International Conference on Massively Parallel Processing*, pp. 142-149, 1998.

[91] M. Ould-Khaoua, "A Performance Model for Duato's Fully Adaptive Routing Algorithm in *k*-Ary *n*-Cubes," *IEEE Trans. on Computers*, vol. 48, no. 12, pp. 1297-1304, 1999.

[92] M. Ould-Khaoua and H. Sarbazi-Azad, "An Analytical Model of Adaptive Wormhole Routing in Hypercubes in the Presence of Hot Spot Traffic," *IEEE Trans. on Parallel and Distributed Systems*, vol. 12, no. 3, pp. 283-292, 2001.

[93] G.J. Pfister and V.A. Norton, "Hot-Spot Contention and Combining in Multistage Interconnection Networks," *IEEE Trans. on Computers*, vol. 34, no. 10, pp. 943-948, 1985.

[94] A. Pombortsis and C. Halatsis, "Performance of Circuit-Switched Interconnection Networks under Nonuniform Traffic Patterns," *Journal of Systems Architecture*, vol. 20, no. 2, pp. 189-201, 1993.

[95] C.s.r. Prabhu, *Grid And Cluster Computing*, Prentice-hall Of India Pvt Ltd 2008.

[96] L. Qiu, Y. Richard, Y.Y. Zhang, and S. Shenker, "On Selfish Routing in Internet-Like Environments," *IEEE/ACM Trans. on Networking*, vol. 14, no. 4, pp. 725-738, 2006.

[97] B. Raman, K. Chebrolu, D. Gokhale, and S. Sen, "On the Feasibility of the Link Abstraction in Wireless Mesh Networks," *IEEE/ACM Trans. on Networking*, vol. 17, no. 2, pp. 528-541, 2009.

[98] S. Ramany, *Routing in Wormhole Networks*, PhD Thesis, Computer Science

Department, University of Saskatchewan, 1995.

[99] A. Raniwala and T. Chiueh, "Architecture and Algorithms for An IEEE 802.11-based Multi-Channel Wireless Mesh Network," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM'05)*, pp. 2223-2234, 2005.

[100] A. Raniwala, D. Pradipta, S. Sharma, R. Krishnan, and T. Chiueh, "End-to-End Flow Fairness Over IEEE 802.11-Based Wireless Mesh Networks," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM'07)*, pp. 2361-2365, 2007.

[101] J. Robinson and E.W. Knightly, "A Performance Study of Deployment Factors in Wireless Mesh Networks," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM'07)*, pp. 2054-2062, 2007.

[102] F. Safaei, A. Khonsari, M. Fathy, and M. Ould-Khaoua, "Pipelined Circuit Switching: Analysis for the Torus with Non-Uniform Traffic," *Journal of Systems Architecture*, vol. 54, no. 1-2, pp. 97-110, 2008.

[103] H. Sarbazi-Azad, M. Ould-Khaoua, and L.M. Mackenzie, "Analytical Modeling of Wormhole-Routed *k*-Ary *n*-Cubes in the Presence of Hot-Spot Traffic," *IEEE Trans. on Computers*, vol. 50, no. 7, pp. 623-634, 2001.

[104] M.D. Schroeder, A.D. Birrell, M. Burrows, H. Murray, R.M. Needham, T.L. Rodeheffer, E.H. Satterthwaite, and C.P. Thacker, "Autonet: A High-Speed, Self-Configuring Local Area Network using Point-to-Point Links," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 8, pp. 1318-1335, 1991.

[105] H.-P. Schwefel, *Performance Analysis of Intermediate Systems Serving*

Aggregated ON/OFF Traffic with Long-Range Dependent Properties, PhD Thesis, Technische Universität München, 2000.

[106] S.L. Scott and G.S. Sohi, "Using Feedback to Control Tree Saturation in Multistage Interconnection Networks," *ACM SIGARCH Computer Architecture News*, vol. 17, no. 3, pp. 167-176, 1989.

[107] S. Shah-Heydari and T. Le-Ngoc, "MMPP Models for Multimedia Traffic," *Telecommunication Systems*, vol. 15, no. 3-4, pp. 273-293, 2000.

[108] H. Sharif and H. Vakilzadian, "Interconnection Network Front-End Controller Combining to Reduce Hot Spots Effects," *Computer Communications*, vol. 20, no. 12, pp. 1089-1097, 1997.

[109] J. Shin and T.M. Pinkston, "The Performance of Routing Algorithms under Bursty Traffic Loads " in *Proc. of International Conference on Parallel and Distributed Processing Techniques and Applications*, pp. 737-743, 2003.

[110] F. Silla, M.P. Malumbres, J. Duato, D. Dai, and D.K. Panda, "Impact of Adaptivity on the Behavior of Networks of Workstations under Bursty Traffic," in *Proc. of IEEE International Conference on Parallel Processing (ICPP'98)*, IEEE Computer Society, Washington, DC, USA, pp. 88-95, 1998.

[111] M.S. Squillante, D.D. Yao, and L. Zhang, "Analysis of Job Arrival Patterns and Parallel Scheduling Performance," *Performance Evaluation*, vol. 36-37, no. 1, pp. 137-163, 1999.

[112] W. Stallings, *Wireless Communications and Networks*, Prentice Hall, 2005.

[113] H. Stefanescu, S. Ciochina, and A. Matei, "Residential Area Service Platforms: Modelling Challenges for Multiple Data Traffic Patterns in Wireless Mesh Networks " in *Proc. of International Conference on Wireless and Mobile Communications (ICWMC'09)*, pp. 340-345, 2009.

[114] T. Vanhatupa, M. Hännikäinen, and T.D. Hämäläinen, "Performance Model for IEEE 802.11s Wireless Mesh Network Deployment Design," *Journal of Parallel and Distributed Computing*, vol. 68, no. 3, pp. 291-305, 2008.

[115] J.S. Vetter and F. Mueller, "Communication Characteristic of Large-Scale Scientific Applications for Contemporary Cluster Architectures," *Journal of Parallel and Distributed Computing*, vol. 63, no. 9, pp. 853-865, 2003.

[116] A. Vishnu, A.R. Mamidala, H.-W. Jin, and D.K. Panda, "Performance Modeling of Subnet Management on Fat Tree InfiniBand Networks using OpenSM," in *Proc. of IEEE International Parallel and Distributed Processing Symposium (IPDPS'05)*, CD-ROM, 2005.

[117] H.J. Wang, B. Raman, C.-N. Chuah, R. Biswas, R. Gummadi, H. Barbara, X. Hong, E. Kiciman, Z. Mao, J.S. Shih, L. Subramanian, B. Y. Zhao, A.D. Joseph, and R.H. Katz, "ICEBERG: An Internet Core Network Architecture for Integrated Communications," *IEEE Personal Communications*, vol. 7, no. 4, pp. 10-19, 2000.

[118] X. Wu, J. Liu, and G. Chen, "Analysis of Bottleneck Delay and Throughput in Wireless Mesh Networks," in *Proc. of IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS'06)*, pp. 765-770, 2006.

[119] Y. Wu, G. Min, K. Li, and A.Y. Al-Dubai, "A Performance Model for Integrated Wireless Mesh Networks and WLANs with Heterogeneous Stations," in *Proc. of Global Communications Conference (GLOBECOM'09)*, 2009.

[120] Y. Wu, G. Min, K. Li, and B. Javadi, "Performance Analysis of Communication Networks in Multi-Cluster Systems under Bursty Traffic with Communication Locality," in *Proc. of Global Communications*

Conference (GLOBECOM'09), 2009.

[121] Y. Wu, G. Min, M. Ould-Khaoua, and H. Yin, "An Analytical Model for Torus Networks in the Presence of Batch Message Arrivals with Hot-spot Destinations," *International Journal of Automation and Computing*, vol. 6, no. 1, pp. 38-47, 2009.

[122] Y. Wu, G. Min, M. Ould-Khaoua, H. Yin, and L. Wang, "Analytical Modelling of Networks in Multicomputer Systems under Bursty and Batch Arrival Traffic," *The Journal of Supercomputing*, in press, online version available at doi:10.1007/s11227-009-0265-x, 2009.

[123] J. Xie and X. Wang, "A Survey of Mobility Management in Hybrid Wireless Mesh Networks," *IEEE Network*, vol. 22, no. 6, pp. 34 - 40, 2008.

[124] Y. Xiong, S. Liu, and P. Sun, "On the Defense of the Distributed Denial of Service Attacks: An On-Off Feedback Control Approach," *IEEE Trans. Systems Man & Cybernetics - Part A: Systems & Humans*, vol. 31, no. 4, pp. 282-293, 2001.

[125] Y. Zhang, Y. Xiao, and H.-H. Chen, "Queueing Analysis for OFDM Subcarrier Allocation in Broadband Wireless Multiservice Networks," *IEEE Trans. on Wireless Communications*, vol. 7, no. 10, pp. 3951-3961, 2008.

[126] Q. Zhao and B.M. Sadler, "A Survey of Dynamic Spectrum Access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79-89, 2007.

[127] J. Zhou, X.-Y. Lin, and Y.-C. Chung, "Hardware Supported Multicast in Fat-Tree-Based InfiniBand Networks," *The Journal of Supercomputing*, vol. 40, no. 3, pp. 333-352, 2007.

[128] J. Zhou and K. Mitchell, "A Single Node Decomposition Based Analytic Model for Multiclass Route Optimization in Wireless Mesh Networks," in

Proc. of IEEE Wireless Communications and Networking Conference (WCNC'07), pp. 3996-4001, 2007.